DOCTOR OF ENGINEERING (ENGD)

Ionospheric Specification and Forecasting

Chartier, Alex

Award date:
2013

Awarding institution:
University of Bath

Ionospheric Specification and Forecasting

Alex Timothy Chartier

A thesis submitted for the degree of Doctor of Engineering

University of Bath

Department of Electronic and Electrical Engineering

November 2013

Summary

Space weather presents a threat to human activities such as Global Navigation Satellite System (GNSS) positioning and timing, power systems, radio communications and transpolar aviation. Nowcasts and forecasts of the ionosphere could help mitigate some of these damaging effects. In this thesis, state-of-the-art ionospheric specification techniques are assessed in a long-term study. That study shows that Global Positioning System (GPS) derived tomographic images specify monthly median ionospheric Total Electron Content (TEC) accurately in Europe and North America throughout the twelve-year test period. Following this assessment, developments are presented in three key areas. The resolution of horizontal structures in ionospheric images over Africa is assessed. The accuracy gains from adding receivers are quantified using a simulation approach, showing that an extended GPS network reduces Root-Mean-Square (RMS) error from 9.5 TEC units for the currently operational network to 4.5 TEC units. A fictional, ideal network is demonstrated to produce images with RMS errors of 3.0 TEC units. Images of the vertical electron density distribution, vital for High Frequency (HF) radio operators, are greatly improved by adding observations of the ionospheric vertical profile to an imaging algorithm that relies on GPS observations. The peak electron density is resolved to an RMS accuracy of 0.5 x 10^{11} electrons/m^3, compared to an RMS accuracy of 1.0 x 10^{11} electrons/m^3 for the standard approach. A novel experimental method is employed to show that forecasts of ionospheric storms could benefit significantly from accurate specification of the initial neutral composition, in particular the ratio of O to N_2 . A theoretical experiment shows that an ideal assimilation of the thermospheric composition can improve storm-time forecasts by at least 10% for over 19 hours, whilst an ideal ionospheric assimilation improves forecasts for less than four hours. This finding will aid the development of a coupled thermosphere ionosphere forecast system.

Acknowledgements

Thank you to my family and friends for their continued support and encouragement. I appreciate the guidance and the freedom given to me by my supervisors, Cathryn Mitchell and David Jackson. Thanks also to the members of the Invert group, past and present, for making it a pleasure to come to work.

I would like to thank the Met Office, the Systems Centre and the Engineering and Physical Sciences Research Council for supporting this work. Thanks to the Jicamarca Radio Observatory, the National Center for Atmospheric Research (NCAR) and Dr. Tomoko Matsuo of the National Oceanographic and Atmospheric Administration (NOAA) for their help through collaborations. I would also like to thank the Software Sustainability Institute for providing me with support through a fellowship.

The work presented in this thesis makes extensive use of data sets provided by the International Global Navigation Satellite System Service (IGS), the United Navstar Corporation (UNAVCO), the African Reference Frame (AFREF), the Low Latitude Ionospheric Sensor Network (LISN), the Brazilian Network for Continuous Monitoring (RBMC), the Jicamarca Radio Observatory (JRO) and the Massachusetts Institute of Technology (MIT) Haystack Observatory. This work also makes use of several models, including the International Reference Ionosphere (IRI) from Union Radio-Scientifique Internationale (URSI) / Committee on Space Research (COSPAR) and the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIEGCM) developed at the NCAR High Altitude Observatory. The M-Map plotting package, developed by Rich Pawlowicz, is used throughout this thesis. In particular, I acknowledge the use of the University of Bath's tomographic imaging algorithm, Multi-Instrument Data Analysis Software (MIDAS).

Table of Contents

List of figures

List of tables

List of abbreviations

Abbreviation	Full name
3DVar	Three-dimensional variational data assimilation
AFREF	African Reference Frame project
AI	Asymmetry Index
ARL-UT	Applied Research Laboratory – University of Texas
ARTIST	Automatic real-time ionogram scaler
CCIR	Comité Consultatif International pour la Radio
CHAMP	Challenging Minisatellite Payload
CME	Coronal Mass Ejection
CODE	Centre for Orbit Detection in Europe
COSPAR	Committee on Space Research
Digisonde	Digital ionosonde
EISCAT	European incoherent scatter radar
eSWua	Electronic Space Weather and Upper Atmosphere
EUV	Extreme Ultra-Violet
F10.7	10.7 cm solar flux
FUSI	Far Ultraviolet Spectrographic Imager
GAIA	Ground-to-topside model of Atmosphere and Ionosphere for Aeronomy
GAIM	Global Assimilation of Ionospheric Measurements
GAIM	Global Assimilative Ionospheric Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HF	High Frequency
hmF2	F2 layer peak height

HWM	Horizontal Wind Model
IDA	Ionospheric Data Assimilation
IDA3D	Ionospheric Data Assimilation Three-Dimensional
IFB	Inter-Frequency Bias
IGS	International GNSS Service
IMF	Interplanetary Magnetic Field
INGV	Istituto Nationale di Geofisica e Vulcanologia
JPL	Jet Propulsion Laboratory
JRO	Jicamarca Radio Observatory
Kp	Planetary magnetic disturbance index
LISN	Low Latitude Ionospheric Sensor Network
MART	Multiplicative Algebraic Reconstruction Tomography
MIDAS	Multi-Instrument Data Analysis Software
MIT	Massachusetts Institute of Technology
MTIEGCM	Magnetosphere Thermosphere Ionosphere General Circulation Model
NCAR	National Center for Atmospheric Research
NIMS	Navy Ionospheric Monitoring System
NmF2	F2 layer peak electron density
NNSS	Navy Navigation Satellite System
OSSE	Observation System Simulation Experiment
PRISM	Parameterised Real-Time Ionospheric Specification Model
RBMC	Brazilian Network for Continuous Monitoring
RMS	Root Mean Square
SUPIM	Sheffield University Plasmasphere Ionosphere Model

TEC	Total Electron Content
TECU	Total Electron Content units / 10^{11} m^{-2}
TGCM	Thermospheric General Circulation Model
TING	Thermosphere Ionosphere Nested Grid
TIGCM	Thermosphere Ionosphere General Circulation Model
TIEGCM	Thermosphere Ionosphere Electrodynamics General Circulation Model
TIMEGCM	Thermosphere Ionosphere Mesosphere Electrodynamics General Circulation Model
TOPEX	Topside Plasmasphere Explorer
UNAVCO	University Navstar Corporation
URSI	Union Radio-Scientifique Internationale
USA	United States of America
USU	Utah State University
USC	University of Southern California
UT	Universal Time
UV	Ultra-Violet

List of symbols

Symbol	Quantity
$\boldsymbol{b}$	Vector of observations
f	Frequency
h	Height (altitude)
k	Signal propagation constant
m	Mass
n	Refractive index
$\boldsymbol{r}$	Vector of solution residuals
s	A point along a ray-path
w	Regularisation weighting
$\boldsymbol{x}$	Vector of model values
A	Observation operator
B	Magnetic Field
D	GPS ionospheric delay term
H	Scale height
I	Radiation intensity
M	Mapping function
N	Density (e.g. electron, ion etc.)
Q	Resolution map
R	Regularisation
T	Temperature (e.g ion, electron etc.)
U	Meridional wind
V	Zonal wind
γ	Non-dispersive GPS signal delay term
ε	Dispersive component of GPS delay on carrier phase
ε_o	Permittivity of free space
ζ	Chapman profile parameter
η	Hessian
λ	Wavelength
σ	Absorption cross section
τ	Dispersive component of GPS delay on pseudo-range

1 Introduction

Since Marconi's transatlantic experiments in 1901, our understanding of the ionosphere has been closely related to radio communications. These experiments prompted Heaviside and Kennelly to suggest that a layer of the upper atmosphere could reflect radio signals and thus allow over-the-horizon communications. Meanwhile, a 1912 Act of Congress in the United States of America (USA) forced amateur radio enthusiasts to use frequencies above 1.5 MHz in the belief that such high frequencies were useless. This legal restriction led to the discovery in 1923 that High-Frequency (HF) radio signal propagation was influenced by the ionosphere. Studies by *Appleton and Barnett* [1925] and *Breit and Tuve* [1925], for which Appleton eventually received a Nobel prize, confirmed the existence of the ionosphere by measuring the angle of arrival and the frequency modulation of reflected radio signals. A summary of contemporary theories of the nature and behaviour of the ionosphere is provided in Chapter 2.

Ionospheric effects on satellite communications have provided both a new means of studying the ionosphere and an additional motivation for doing so. The ionosphere reduces the speed of signals passing through it by a frequency-dependent amount, which means a dual-frequency ground-based receiver can measure the density of electrons between it and the satellite. In a pioneering study, *Austen et al.* [1986] applied tomographic techniques, which were initially developed for medical applications, to the problem of imaging the ionosphere. Signals received on the ground from satellites at orbits of several hundred kilometres took the place of X-rays travelling through the human body. These initial efforts employed chains of receivers arranged along the orbit trajectory of a single satellite and therefore produced images of a two-dimensional slice of the ionosphere. More recently, ionospheric imaging has been extended to three dimensions by using multiple satellites and globally distributed receivers from Global Navigation Satellite Systems (GNSS), as well as a range of other ionospheric observations. The history and current state of ionospheric imaging is described in Chapter 3.

Ionospheric effects on GNSS systems, such as the Global Positioning System (GPS),

include positioning errors of up to 100 m, according to *Klobuchar* [1987]. These errors can be mitigated by including an ionospheric correction to the positioning estimate. Whilst GNSS-based ionospheric tomography can provide accurate corrections where there is good observation coverage, poorly-observed regions such as over the oceans and the polar caps require a different approach. Empirical models such as the International Reference Ionosphere (IRI) rely on long-term, statistical averages of ionospheric observations to estimate current conditions based on solar and geomagnetic parameters. The study presented in Chapter 4 assesses the accuracy of both ionospheric tomography and the IRI empirical model in specifying the ionosphere over an entire twelve-year solar cycle and across three different geographic regions.

Following this baseline study, the limitations of ionospheric imaging are addressed in three key areas: horizontally, vertically and in time. Difficulties in resolving the horizontal structure of the ionosphere are assessed in Chapter 5. Africa is chosen as the location of study here, because the relative lack of observations makes ionospheric imaging more challenging. A simulation approach demonstrates that improved imaging accuracy can be achieved by including additional GPS receivers. Accurate imaging of the vertical structure, which is especially important for High-Frequency (HF) radio communications, is addressed in Chapter 6. The vertical profile is poorly constrained in GPS-derived ionospheric images because most of the rays arrive at relatively high angles, especially at low latitudes. The powerful Jicamarca incoherent scatter radar, located near the magnetic equator in Peru, is used to validate the improved specification of the vertical profile presented in this chapter. In Chapter 7, the problem of ionospheric forecasting is investigated from the perspective of designing a successful future system. The goal of this study is to find out which parts of the ionosphere and the thermosphere at the beginning of a storm have the most bearing on the way the storm plays out. An additional aim of this study is to compare the potential forecast accuracy improvements that can be obtained by knowing the initial thermosphere and ionosphere against the benefits of knowing the external solar and geomagnetic drivers.

2 Physics of the ionosphere

Abstract

In this chapter, an overview of ionospheric physics is given to put the results shown in later chapters into context. The important physical processes surrounding production, loss and transport of ionospheric plasma are explained. The established physics described in this chapter is drawn from textbooks by *Hargreaves* [1979], *Rishbeth and Garriott* [1969], *Kivelson and Russell* [1995] and *Kelley* [2009].

2.1 Introduction

The ionosphere is a region of the upper atmosphere that contains sufficient free charges to influence the propagation of radio waves. The ionosphere is created primarily by the photo-ionization of neutral particles in the upper atmosphere. *Rishbeth* [1988] suggests that the ionosphere can be considered to exist between 60 and 600 km altitude, at least for practical purposes.

The ionosphere is scientifically very useful. Its easily observable nature allows it to be used as a tracer for studying the upper atmosphere [*Rishbeth,* 1988]. However it should be noted that, even at the peak of ionization, less than 1% of particles are ionized. Additionally, the ionosphere is electrically neutral to a high degree of approximation. This neutrality occurs because positive and negative charges are always created in pairs, by the processes of ionization and recombination. Even under the influence of a strong electric field, the charge imbalance remains a minute fraction of the total charge density. Charge neutrality is maintained because of the strong electrical attraction between positive and negative particles. The small charge imbalances that occur can lead to electric potentials of 250 kV above ground.

This review begins with a description of the neutral atmosphere, which provides the particles that become ionized to form the ionosphere. Within the ionospheric section, the local photochemistry and energetics of the ionosphere will be discussed first, followed by an explanation of plasma transport processes and an overview of

ionospheric electrodynamics.

2.2 Neutral Atmosphere

The neutral atmosphere is composed of a number of chemical species, predominantly molecular oxygen, O_2 , and molecular nitrogen, N_2 . The different species are kept in equal proportions below around 100 km altitude by the process of turbulent mixing. Above this height, molecular diffusion dominates, meaning that the different chemical species are stratified into layers. Each species obeys its own partial form of the ideal gas law, forming layers defined by the temperature of the atmosphere and the mass of the species. Atmospheric temperatures reach a minimum of around 130 K in the mesosphere, between 85 – 100 km altitude. Above this altitude, temperatures in the thermosphere can reach over 2000 K. Despite being the most abundant species in the atmosphere, O_2 and N_2 are dominant only in the lower part of the thermosphere, below 200 km. This is because O_2 and N_2 have higher masses than the other constituents of the upper atmosphere: H, He, O and N. Above 200 km, the thermosphere is dominated by atomic oxygen, O. The transition from O to H, at about 600 km, marks the end of the thermosphere and the beginning of the exosphere. The different atmospheric layers are shown in Figure 1. *Chapman* [1931] proposed that the ionospheric layers (shown alongside the atmospheric layers in Figure 1) are formed by ionization of the neutral gas by ultraviolet and X-ray sunlight. Different wavelengths of light ionize the different neutral species, so the combination of the depth of penetration of a given wavelength of light into the atmosphere and the composition of the neutral atmosphere dictate the rate of ionization, or plasma production. Ionization of molecular oxygen is responsible for the E layer, while ionization of atomic oxygen and molecular nitrogen creates the F layer. The resulting plasma is distributed according to diffusive equilibrium along the Earth's magnetic field lines and recombines faster at lower altitudes where the air is denser. The weakly ionized D layer, which exists from 60 – 90 km, and the daytime splitting of the F layer into the constituent F1 and F2 peaks have little bearing on ionospheric imaging since their existence has little bearing on ionospheric tomography.

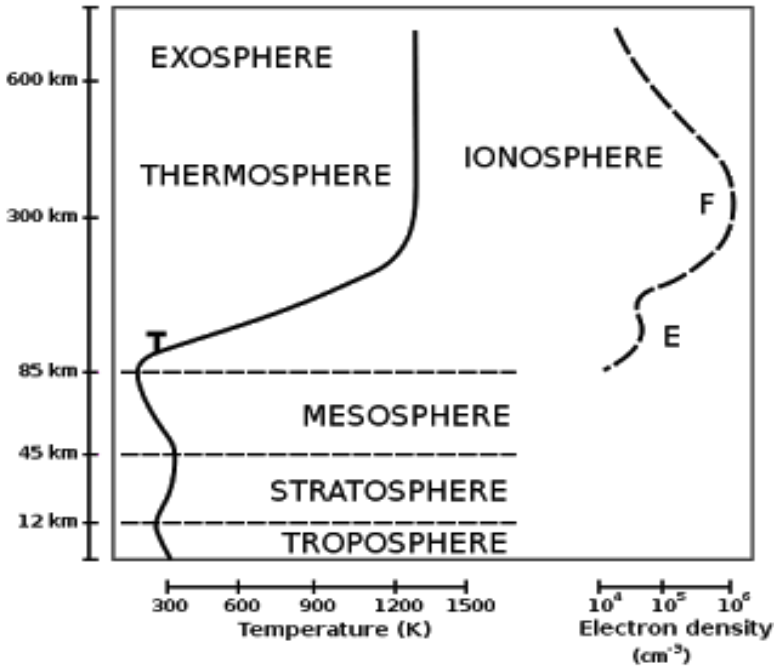

Figure 1: Temperatures through the atmosphere, left, and ionospheric electron densities, right. Retrieved from http://en.wikipedia.org/wiki/File:Atmosphere_with_Ionosphere.svg on 1 November 2013

2.3 Photochemistry

Absorption of Extreme Ultra Violet (EUV) radiation by thermospheric neutral species leads to photo-ionization, which creates the bulk of the plasma that makes up the ionosphere. Equation (1) describes the rate of photo-ionization for one ion species, q_i , where s is a point along the path of the light, I is the intensity of light, λ is the wavelength, N_n is the density of neutral species n and σ_n^{ion} is the absorption cross section of neutral species n.

$$q_i(s) = \int d\lambda \ \Sigma \ \sigma_n^{ion}(\lambda) \ N_n(s) \ I(s, \lambda) \qquad (1)$$

The attenuation of EUV flux, I, along the path of light is described by Lambert's law:

$$dI / ds = - \Sigma \ \sigma_n^{abs}(\lambda) \ N_n(s) \ I(s, \lambda) \qquad (2)$$

Combination of the typical Solar EUV flux and the distribution of neutral species leads to a peak in O_2^+ production around 100 km, a peak in N_2^+ production around 150 km and a peak in O^+ production around 160 km.

Solar EUV photons ($h\nu$) with energies above 13.6 eV ionize atomic oxygen as follows:

$$O + h\nu \longrightarrow O^+ + e^- \tag{3}$$

The most straightforward recombination process is radiative recombination, which is the reverse of Equation (3) above:

$$O^+ + e^- \longrightarrow O + h\nu \tag{4}$$

Radiative recombination is very slow, so most ions recombine by atom-ion charge exchange followed by dissociative recombination. For the case of O^+, two common forms of atom-ion charge exchange are shown below:

$$O^+ + O_2 \longrightarrow O_2^+ + O \tag{5}$$

$$O^+ + N_2 \longrightarrow NO^+ + N \tag{6}$$

Equations (5) and (6) show two mechanisms for the creation of molecular ions from atomic ions and molecules. These molecular ions can undergo dissociative recombination, so long as free electrons are present:

$$NO^+ + e^- \longrightarrow N + O \tag{7}$$

$$O_2^+ + e^- \longrightarrow O + O \tag{8}$$

Dissociative recombination reactions, such as the ones shown in (7) and (8), are much faster than radiative recombination reactions because they do not require the emission of a photon. However, a photon was initially absorbed to ionize the particles on the left of (7) and (8), so one of the resulting atoms must be in an excited state. This will eventually be released and create airglow.

In the case of atomic oxygen, the reaction can be:

$$O\,(^1D) \longrightarrow O\,(^3P) + h\nu\,(630.0\text{ nm}) \tag{9}$$

$$O\,(^1S) \longrightarrow O\,(^1D) + h\nu\,(557.7\text{ nm}) \tag{10}$$

Photons emitted in reactions (9) and (10) have characteristic wavelengths that depend on the type of species that has been excited and the level of excitation. Many of the emissions are in the visible spectrum - atomic oxygen emits green and red

light, whilst atomic nitrogen emits blue and red light. When excited by particle precipitation, these emissions can form the only ionospheric phenomenon visible to the naked eye – the Aurora.

2.4 Energetics

In any gas or plasma, energy can be transferred between particles by collisions. In the case of elastic collisions, the total kinetic energy of the particles remains the same before and after a collision. The mass ratio of two colliding particles dictates how their velocities will be affected by the collision. If a light, moving particle impacts a heavy, stationary particle, both will continue moving at about the same speed. However, if a heavy, moving particle impacts a light, stationary particle, the lighter particle will attain a very high velocity. Ions are far more massive than electrons, so collisions between ions and electrons have the effect of increasing electron velocities without having much effect on ion velocities. The result is that electron temperatures are often much higher than ion temperatures.

In the case of inelastic collisions, the total kinetic energy of the particles is not the same before and after a collision. Inelastic collisions are common when molecules are involved, because energy can be transferred from the motion of a molecule into the molecule's internal degrees of freedom. In the case of a collision between an electron and molecular nitrogen, the following process may occur:

$$N_2 + e^- \longrightarrow N_2(vib) + e^- \tag{11}$$

Equation (12) shows how vibrationally excited N_2 can be created by collision with electrons. *Richards and Torr* [1986] showed that enhanced levels of vibrationally excited N_2 in the thermosphere, as occurs in summer at solar maximum, significantly elevate the rate of the $O^+ + N_2$ reaction shown in Equation (6). This can lead to a a factor of two decrease in the peak solar maximum O^+ density, compared to the level that would be expected without the vibrationally excited N_2 mechanism.

Collisions between electrons and atoms may also be inelastic. In Equation (12), atomic oxygen takes on electronic fine structure through collision with an electron.

$$O\,(^3P) + e^- \longrightarrow O\,(^3P^*) + e^- \quad (12)$$

Mlynczak et al. [2004] show observations of the atomic oxygen fine structure emission line in the upper mesosphere and lower thermosphere. Inelastic collisions are also responsible for the quenching process that acts to prevent certain characteristic emissions. Processes such as those shown in Equations (13) and (14) allow excited neutral species to be deactivated without releasing photons:

$$O\,(^1D) + N_2 \longrightarrow O\,(^3P) + N_2\,(vib) \quad (13)$$

$$N\,(^2D) + e^- \longrightarrow N\,(^4S) + e^{-*} \quad (14)$$

These reactions act to preserve energy in the thermosphere-ionosphere system rather than have it radiate away in the form of photons. In the case of Equation (14), the asterisk indicates that the free electron has taken on energy from the nitrogen atom.

2.5 Transport and Electric Fields

Three forms of pressure act on the ionosphere: dynamic, thermal and magnetic. Dynamic pressure is proportional to the density of the fluid multiplied by the square of its velocity. This type of pressure is predominantly caused by thermospheric winds. Thermal pressure is proportional to the number density of the fluid multiplied by its temperature, which is defined by the processes described in Section 2.4. Magnetic pressure is proportional to the square of the magnetic field strength. The magnetic pressure exerted on ionospheric particles is far stronger than dynamic and thermal pressure, so transport parallel and perpendicular to the magnetic field are fundamentally different. Details of the Earth's magnetic field will be given in Section 2.6. Additionally, it should be noted that the ionosphere always remains neutral or nearly neutral over macroscopic volumes – electric attraction prevents the development of a net flow of electrons away from ions.

The important transport processes are caused by externally imposed electric fields, neutral winds, thermal expansion and contraction, and diffusion. Since transport processes are so tied up with electric fields and currents, the two issues will be

discussed together.

Ionospheric electric currents can arise from electric fields. The rate of current flow depends on the electrical conductivity and mobility of the charged particles present. Ionospheric electric fields have a number of different origins. Sometimes, ionospheric electric fields arise from current flows caused by other internal processes, which will be discussed below. However, ionospheric electric fields can also be imposed externally, by the magnetosphere. This takes the form of a potential difference across a region of the ionosphere. When an electric field exists perpendicular to a magnetic field, a force in the ExB direction is created. This process, known as ExB drift, is described by Fleming's right hand rule.

When a conductor is forced to move across a magnetic field, a current is generated. The ionosphere can be forced across the Earth's magnetic field lines by thermospheric neutral winds, generating currents. Charged particles will experience a force proportional to the difference between the wind velocity and the velocity of the charged particles, and to the collision frequencies. The ionospheric E region is exposed to mainly tidal winds caused by thermal expansion and contraction of the atmosphere. The action of thermospheric tides on the ionosphere is responsible for the atmospheric dynamo effect. This effect sets up the electric fields responsible for the equatorial ionization anomaly. Temperature changes affect charged particles in the same way as neutral particles, so the ionosphere goes through its own daily thermal expansion and contraction. The ionospheric plasma diffuses in the same way as a gas, under the influence of gravity and of gradients in its own partial pressure. Electrostatic attractions keep the ions and electrons together unless strong electric fields are present. Collisions with neutral particles restrict the speed of diffusion in the lower ionosphere, but have little effect in the F region and above. Earth's magnetic field constrains the direction of diffusion so that it only occurs in the direction of the field lines.

2.6 Earth's magnetic field

The Earth's magnetic field can be approximated to a magnetic dipole angled around 11 degrees from the axis of rotation, though it should be noted that the field deviates

significantly from dipolar geometry in the American longitude sector. The Earth's magnetic field is made up of open and closed magnetic field lines. The open field lines meet the Earth at high latitudes whilst the closed field lines enter the Earth at high and middle latitudes. Open field lines connect with the solar wind, night-side field lines extend huge distances back into the magnetotail on the Earth's night-side, whilst the closed field lines remain far closer to the Earth. The boundaries between open and closed field lines are called the cusps. The cusp regions are important because they define the locations of the auroral ovals. These narrow gaps, about one degree in latitude wide at the surface, allow solar wind plasma to penetrate into the magnetosphere and down into the high-latitude ionosphere. These fast-moving particles impact the high-latitude thermosphere, creating highly structured regions of ionization and occasionally exciting the neutral species enough to create the aurorae. Some common mechanisms for the radiation of visible light were discussed in Section 2.4.

2.7 Summary

In this chapter, it was shown that the ionosphere is a region of the Earth's upper atmosphere that is created primarily by photo-ionization of neutral particles in the thermosphere. The chemistry and vibrational excitation of the thermosphere have important effects on the rates of recombination in the ionosphere. The ionospheric plasma is moved around by the combination of dynamic, thermal and magnetic pressure. The Earth's magnetic field interacts with electric fields in the plasma to produce drifts in the ExB direction. In Chapter 3, the ionosphere's effect on radio signals will be explained both as a motivation for and a means to study the ionosphere. Methods that use the physics described in this chapter to improve ionospheric specification and forecasting will be put forward throughout this thesis.

3 Ionospheric imaging

Abstract

The history of ionospheric imaging is described in this chapter. Ground-based GPS receivers now provide the main source of ionospheric observations for tomography, although low Earth orbit satellites historically provided the measurements for this purpose. More recently, data assimilation approaches have been developed for ionospheric specification, some of which include a modelled thermospheric component.

3.1 Introduction

The field of ionospheric imaging has grown up largely from the perspective of tomography. *Bust and Mitchell* [2008] give a review of ionospheric imaging techniques, focussing on techniques that use observations from GPS and dual-frequency beacon transmitters, as well as in-situ electron density measurements.

3.2 Ionospheric observations

Ionospheric observations can be divided into measurements of electron density and measurements of Total Electron Content (TEC), which is an integrated measure of electron density along a path traced through the ionosphere. In ionospheric applications, two types of TEC are commonly referred to: slant TEC, which is the number of electrons along the path between a satellite and a receiver, and vertical TEC, which is the number of electrons in a column from the ground up to an arbitrary height. Slant TEC is observed using ground- and space-based GNSS receivers, as well as by low Earth orbit Beacon transmitters. Vertical TEC can be observed by satellite altimeters, such as the Topside Plasmasphere Explorer (TOPEX) instrument aboard the Jason satellites, but is generally an output product of

ionospheric specification algorithms. Electron density measurements come from ground-based instruments such as ionosondes, incoherent scatter radars and in-situ instruments aboard low Earth orbit satellites such as the Challenging Minisatellite Payload (CHAMP). GPS-derived observations of slant TEC are described here because they are used throughout this thesis, whilst ionosonde and incoherent scatter radar observations are dealt with in Sections 6.1.3 and 6.1.4.

3.2.1 GPS slant TEC observations

Signals transmitted from GPS satellites to receivers, either on the ground or on low-earth orbit satellites, experience a delay due to the ionosphere. The ionospheric delay is directly related to the electron density along the signal's path. The total integrated electron density along the path is referred to as slant TEC. Example GPS slant TEC measurements are illustrated in Figure 2.

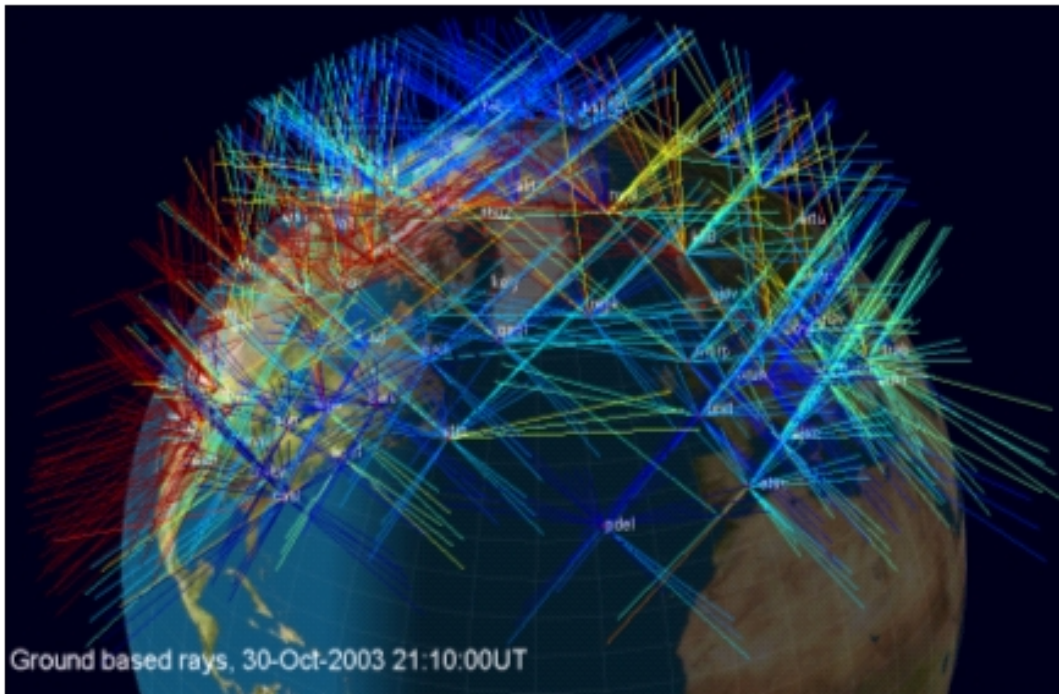

Figure 2: Slant TEC measured by dual-frequency GPS receivers on 30 October 2003. Reproduced from Mitchell et al. *[2008] with permission.*

Ionospheric delays cause positioning errors for single-frequency GPS users, but can be removed by dual-frequency receivers. *Davies* [1990] explains how the ionosphere affects radio signal propagation, solving for the group refractive index, n_g, from the Appleton-Hartree equation. The assumption used throughout is that the ionosphere

is a cold plasma in a weak magnetic field. The Appleton-Hartree equation, shown in Equation 15, solves for the phase refractive index, n_p, from the density of free electrons, N_e, the relationship between the charge of an electron, e, and its mass, m_e, as well as the frequency of the signal, f.

$$n_p^2 = 1 - N_e e^2 (4 \pi^2 \varepsilon_o m_e f^2)^{-1} \quad (15)$$

From this equation, a constant, k, can be extracted to collect all the natural constants:

$$n_p^2 = 1 - k N_e f^{-2} \quad (16)$$

Here, k is about 80.6 $m^3 s^{-2}$. *Davies* [1990] shows that the group refractive index, n_g, is related to the phase refractive index, n_p, as follows.

$$n_g = \frac{d}{df} (f n_p) \quad (17)$$

Using a binomial expansion to the first order, the relationship in Equation 16 can be approximated as follows:

$$n_p \approx 1 - 1/2\ k N_e f^{-2} \quad (18)$$

By combining equations 17 and 18, we can find the group refractive index, n_g, in terms of electron density, N_e, and signal frequency, f.

$$n_g \approx 1 + 1/2\ k N_e f^{-2} \quad (19)$$

Mannucci et al. [1999] use equations 18 and 19 to describe the ionospheric delay terms in the four observables produced from dual-frequency GPS signals. Equations 20 – 23 show the pseudo-range (group) delays, P_1 and P_2, and the carrier phase advances, L_1 and L_2. f_1 and f_2 are the frequencies of the two signals, while D is the ionospheric delay term. Each observable includes a non-dispersive term, ɣ, that combines the geometric distance, tropospheric delays, clock errors and non-dispersive hardware delays. The carrier phase terms have an integer cycle ambiguity ($\lambda_1 n_1$, $\lambda_2 n_2$) where λ_1 and λ_2 are the signal wavelengths and n_1 and n_2 are unknown integers. Cycle slips can occur when lock is lost by the receiver, but these can be removed using a detection algorithm described by *Blewitt* [1990]. ε and τ are the dispersive components of the satellite and receiver hardware biases.

$$P_1 = \gamma + 40.3\ D f_1^{-2} + \tau_1^r + \tau_1^s \quad (20)$$

$$P_2 = \gamma + 40.3\ D f_2^{-2} + \tau_2^r + \tau_2^s \quad (21)$$

$$L_1 = \gamma - 40.3\ D f_1^{-2} + \lambda_1 n_1 + \varepsilon_1^r + \varepsilon_1^s \quad (22)$$

$$L_2 = \gamma - 40.3\ D f_2^{-2} + \lambda_2 n_2 + \varepsilon_2^r + \varepsilon_2^s \quad (23)$$

Frequency-differencing the two pairs of equations removes the non-dispersive components of the delay or advance, leaving the ionospheric contribution and the inter-frequency biases. The inter-frequency biases are the combined, differenced satellite and receiver biases, ε and τ, and are referred to as IFB for the pseudo-range and IFB' for the carrier phase.

$$P_I = P_2 - P_1 = 40.3\ D\ (f_2^{-2} - f_1^{-2}) + \text{IFB} \quad (24)$$

$$L_I = L_2 - L_1 = 40.3\ D\ (f_2^{-2} - f_1^{-2}) + (\lambda_2 n_2 - \lambda_1 n_1) + \text{IFB}' \quad (25)$$

These inter-frequency biases, or IFBs, change over time [*Wilson and Mannucci,* 1993; *Sardon and Zarraoa, 1997*], but can be solved for during the ionospheric imaging process, as is shown by *Mannucci et al.* [1998]. It is possible to remove IFBs from the observations altogether by taking the difference between measurements from the same satellite-receiver pair at different times. The 'observation' is then a difference in slant TEC between two rays from different times and locations. It is preferable to use observations derived from carrier phase for this process because they have far less noise than pseudo-range observations. If the carrier phase is used, phase lock must be maintained in the intervening period between the two measurements so that the integer cycle ambiguity remains constant.

3.3 Ionospheric Tomography

Two- and three-dimensional tomographic techniques exist for imaging the ionosphere. *Radon* [1917] was the first to solve the problem of reconstructing a function from its projections. The first application of tomography was published by *Bracewell* [1956], who used a 'fan beam' from a radio telescope to image celestial light distributions. Computer aided tomography was invented by *Hounsfield* [1972], who reconstructed images of the human body from measurements of the attenuation of radiation passed through it at varying angles. An example three-dimensional

tomographic image of electron density is shown in Figure 3:

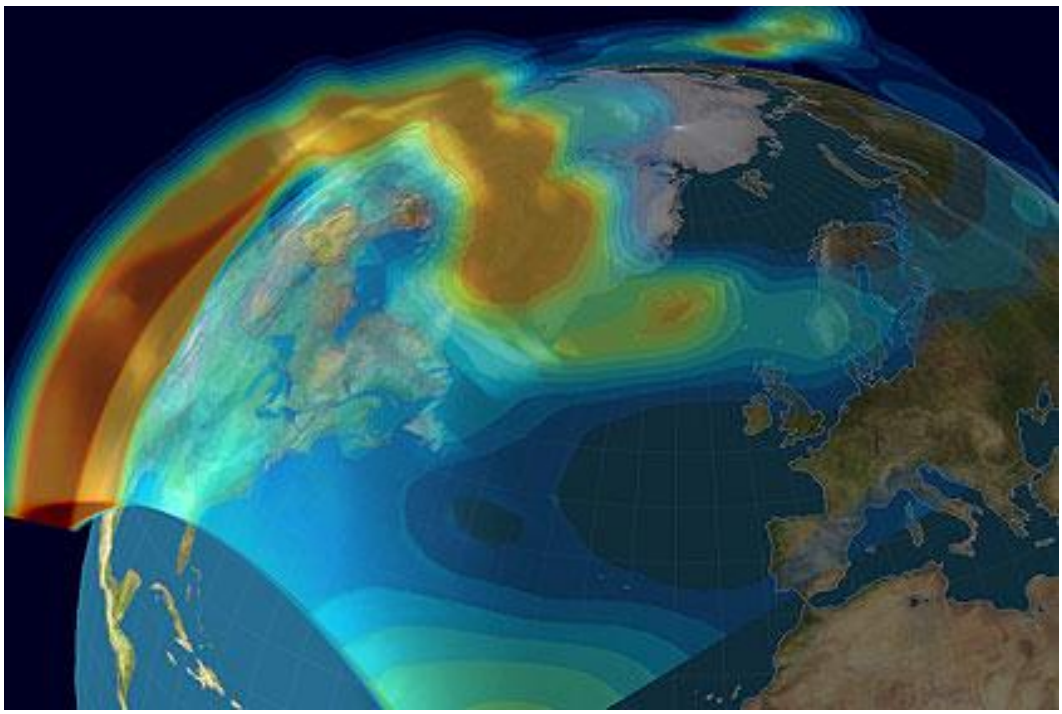

Figure 3: An image of the Halloween 2003 ionospheric storm produced by the MIDAS tomographic imaging algorithm. Reproduced from Mitchell et al. *[2008] with permission.*

Computerised ionospheric tomography was first demonstrated by *Austen et al.* [1986, 1988] in a simulation study for two-dimensional imaging. The authors used a grid of varying latitude and height at a fixed longitude, with pseudo-observations of slant TEC derived from their simulated ionosphere in positions that would have been observable by the Navy Navigation Satellite System (NNSS) satellites. This gave them an array of rays in a fan-like distribution. The authors used an iterative finite series expansion algorithm, as described by *Censor* [1983], since that method allowed any ray-path geometry to be used. This method is based on discretizing the image domain before any mathematical analysis takes place, whereas transform methods only discretize their formulas at the very end. *Raymund et al.* [1990] presented another iterative method, the Multiplicative Algebraic Reconstruction Technique (MART), for reconstructing a vertical slice of electron densities. MART is an entropy-optimization algorithm developed by *Gordon et al.* [1970], which is considered suitable for situations with incomplete observations. In the work of *Raymund et al.* [1990], an initial guess for the electron densities is achieved by dividing the observed slant TEC along a path into the pixels according to the path lengths within each pixel. This is combined with a weighting based on Chapman

profiles, as was described by *Hargreaves* [1995]. The method produced good agreement between the reconstructions and the original simulated images.

Some of the theoretical limitations of ionospheric tomography were investigated by *Yeh and Raymund* [1991]. The authors found that, in the absence of continuous data from all angles or an idealized plane geometry, it would not be possible to find an exact and unique solution. The authors noted that the spherical shape of the Earth meant rays close to horizontal are not available through the ionosphere, which restricts the vertical resolution that is achievable with ionospheric tomography. Refraction of low elevation rays also degrades the quality of those observations which are closest to horizontal. The discrete locations of the receivers pose a further limitation to the accuracy of the images. The variation of the ionosphere over the data collection period breaks the assumption of a static ionosphere, which prompted the authors to suggest a time-varying algorithm, such as the one later developed by *Mitchell and Spencer* [2003].

Raymund et al. [1994] used a least-squares technique to produce ionospheric images by linear combination of model ionospheres. The authors found the inclusion of scaled ionosonde data was of particular benefit, but the technique also allowed for the use of both relative and absolute slant TEC observations. However, none of the model ionospheres contained small-scale structures, so it was not possible to produce detailed images. *Raymund et al.* [1993] found they could introduce these small-scale structures by first producing a smooth background ionosphere by linear combination of model ionospheres, then using that as the basis for an iterative algorithm to fill in the detail. It should be noted that, whilst this modelled ionosphere is described as a 'background', it has a different function to the model background used in meteorological variational data assimilation techniques. In the case of three-dimensional varational data assimilation (3DVar), for example, the solution is effectively a weighted average between the observations and the model background, whereas *Raymund et al.'s* [1993] ionospheric background is simply a starting point for the iterative process, which finds a least-squares fit to the observations only.

Na and Lee [1990] described a technique for two-dimensional electron density imaging using Fourier orthonormal basis functions based on prior information. *Sutton and Na* [1994] showed the technique produced promising results using

simulation data, but explained that the reconstructed vertical profile depended strongly on prior information. *Cornely* [2003] explained that this is a consequence of insufficient data coverage.

Fremouw et al. [1992] used vertical orthonormal functions, derived from an ionospheric model, and horizontal functions, derived from a power law spectrum, to image ionospheric electron densities. *Fougere* [1995] instead used a linear combination of orthonormal functions derived from Chapman profiles. An example of this process is shown in Figure 4:

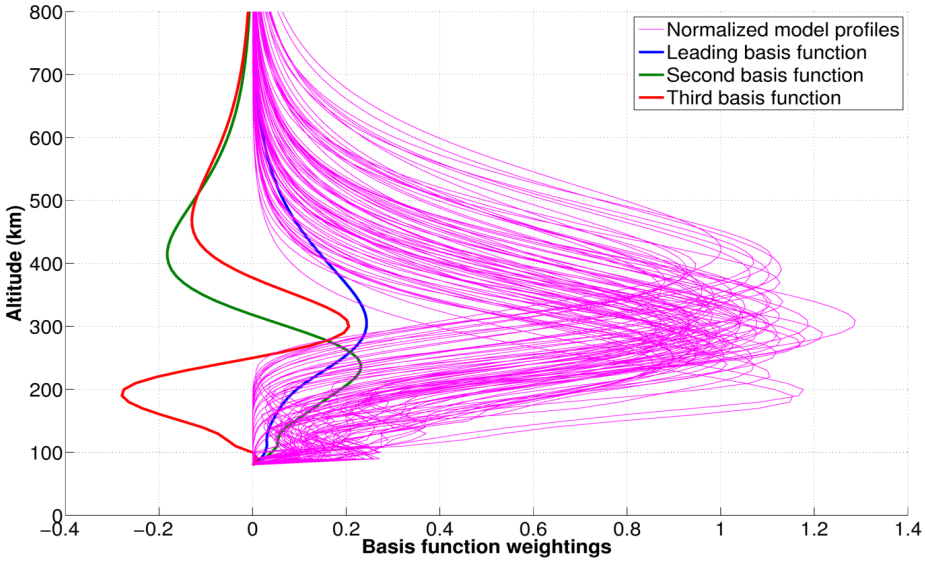

Figure 4: Three orthonormal functions (red, green and blue) derived from Chapman profiles (purple) using singular value decomposition.

Kuklinski [1997] created a three-dimensional imaging grid of size 0.5° x 0.5° x 45 km. Slant TEC observations were used to create tomographic images of a plane defined by a chain of receivers and a NNSS satellite orbit path. *Kuklinski* [1997] also presented simulation results from a network of over 50 receivers spread over the United States of America (USA). This highlighted the possibility of performing quasi-three-dimensional images of the ionosphere from a single NNSS pass, an idea explored further by *Mitchell et al.* [1997a]. *Na and Sutton* [1994] explored the resolution limits of ionospheric tomography and found that, although the theoretical resolution improved as the number of receivers was increased, there was a limit of around 50 km horizontal resolution that could be achieved using tomography. *Heaton et al.* [1995] achieved some improvements to the vertical resolution by including ionosondes in the inversions. *Mitchell et al.* [1997b] showed that an

inversion constrained with Chapman profiles could retrieve the peak height to within about 30 km, but that the relationship between peak density and scale height was ambiguous. For example, the inversion could underestimate peak density and compensate by overestimating scale height. The research presented in Chapter 6 of this thesis addresses the issue of resolving peak density and scale height by including ionosondes in the inversion algorithm.

3.4 Three-dimensional Imaging Theory

Hajj et al. [1994] highlighted the potential to use ground-based and space-based GPS observations to image the ionosphere. The authors suggested that satellite-to-satellite signals from GPS transmitters to receivers in low Earth orbit could specify the vertical structure of the ionosphere, whilst ground-based receivers of NNSS satellite transmissions could identify the horizontal structures. *Rius et al.* [1997] showed the first experimental results of ionospheric imaging using GPS data. The authors applied a Kalman filter approach to the problem of ionospheric imaging. *Howe et al.* [1998] also applied the Kalman filter approach to ionospheric tomography in order to create a full three-dimensional ionospheric imaging algorithm.

The sparse coverage of ionospheric observations means that assumptions have to be made in order to produce three-dimensional images of the ionosphere. One approach, employed by *Howe et al.* [1998], is to transform a problem of solving for electron densities into one of solving for basis function coefficients. Another approach is to choose Sun-fixed coordinates, exploiting the solar EUV dominance of the ionization to help fill data gaps [*Juan et al.,* 1997]. However, this is not necessarily the best frame for imaging the disturbed ionosphere, which is not dominated by changes in insolation.

A simple inversion problem is to solve the equation:

$$\boldsymbol{b} = A\,\boldsymbol{x} \tag{26}$$

where $\boldsymbol{b}$ is a vector of slant TEC observations and A is a matrix of ray-path geometry (an observation operator) that transforms the electron density solution, $\boldsymbol{x}$, to the

form and location of the observations.

There are different ways of using GPS slant TEC measurements. One method is to take the differential pseudorange estimate of slant TEC, smooth it using the differential phase estimate and then correct for the satellite and receiver interfrequency biases using pre-calculated estimates. The estimation of these biases is a source of uncertainty in methods that use calibrated observations. *Ciraolo et al.* [2007] showed that these biases can be much larger than the few TEC units that were originally estimated. Another method for observing TEC is to use the TEC changes from a continuous satellite-receiver path. This way, any fixed bias in the observations is removed – only the relative difference is necessary. This method is achieved by differencing the appropriate lines of the geometry matrix, A, and the observation vector, *b*. The 2-D inversion algorithms of *Andreeva* [1990] utilized this method. Of course, inversions that use this type of observation must at least implicitly calculate their own calibration constants.

Unfortunately, Equation (26) cannot by itself be used to solve for electron density in any real problem because the observation geometry is not suitable. The problem is under-determined. This means that multiple solutions exist, each of which satisfies the available information. Several approaches exist for overcoming this problem. One method is to introduce prior assumptions to the algorithm in order to require physically or empirically acceptable solutions. This can be achieved by reducing the problem of solving for electron densities to one of solving for basis function coefficients. These basis functions are formed from empirical or physics-based models and provide strong constraints to the behaviour of the solution in any given dimension. The basis function approach has been used by *Fremouw et al.* [1992] in two dimensions, by *Howe et al.* [1998] in three dimensions using a Kalman filter and by *Mitchell and Spencer* [2003] in a four-dimensional algorithm. In some cases, basis functions are tailored to the specific conditions (local time, date, location, solar and geomagnetic activity etc.), as is described in *Materassi and Mitchell* [2005a, 2005b]. Horizontal spherical harmonics may also be truncated in order to minimize aliasing in data-sparse regions [*Mitchell and Spencer,* 2003].

3.5 Data assimilation approaches to ionospheric specification

There is a separation between techniques designed simply to specify the ionosphere at an instant in time, such as is the case in tomography, and techniques aimed at correcting the path of a model as it runs forwards in time. Data assimilation techniques start from a model state estimate, called a *background*. The background is combined with observations to make an optimal estimate of the state, which is called an *analysis*. The analysis can then be used as the starting point for another forecast, and the process is repeated. One advantage of data assimilation over tomographic imaging is that the use of a background helps to overcome the lack of ionospheric observations in regions such as the oceans. However, there is a risk that a poor model could introduce biases into the analysis. Current ionospheric data assimilation approaches only 'forecast' up to the present time, at which point they add more observations, but data assimilation could potentially form part of a system capable of forecasting several hours into the future. Research aimed at defining the underpinnings of a future ionospheric forecast system is presented in Chapter 7.

Several groups have taken an approach to ionospheric specification. An algorithm known as Ionospheric Data Assimilation Three-Dimensional (IDA3D) was developed at the Applied Research Laboratory at the University of Texas (ARL-UT) by *Bust et al.* [2000, 2004]. IDA3D is a 3DVar technique, so it relies on observation and background error covariance estimates to combine the background model (typically either IRI-2001 by *Bilitza* [2001] or the Thermosphere-Ionosphere-Mesosphere-Electrodynamics General Circulation Model (TIMEGCM) by *Roble and Ridley* [1994]) with the observations to form a state estimate or analysis. In addition, IDA3D uses the Gauss-Markov Kalman filter technique, described by *Gelb* [1974] to propagate the analysis and its error covariance forward in time.

Utah State University (USU) has also developed an ionospheric data assimilation scheme, called Global Assimilation of Ionospheric Measurements (GAIM) [Schunk et al., 2004, Schunk et al., 2005a, 2005b; Scherliess et al., 2004]. One version of USU GAIM, described by *Schunk et al.* [2004], is mathematically identical to IDA3D, but uses the Ionospheric Forecast Model [*Schunk et al.*, 1997] as its background model. Another version of USU GAIM, described by *Scherliess et al.* [2004], utilizes a

reduced-state Kalman filter. The background error covariance is calculated in a reduced-state space, where the number of variables is reduced by a factor of 100, increasing computational efficiency by a factor of 10 000. In addition, the state and error covariance estimates are evolved forwards in time by a linearized version of the model. This means that the full, non-linear model does not need to be run multiple times, as would be the case with an ensemble Kalman filter.

Another, separate assimilation scheme called GAIM exists. This one is the Global Assimilative Ionospheric Model, developed by the University of Southern California (USC) and the Jet Propulsion Laboratory (JPL) [*Wang et al.*, 2004; *Hajj et al.*, 2004; *Pi et al.*, 2003]. USC/JPL GAIM uses a simplified version of the Sheffield University Plasmasphere Ionosphere Model (SUPIM), presented by *Bailey et al.* [1997]. The simplified version of the model relies on a single ion species, O^+. This ion density model is driven by various empirical models.

3.6 Results and validation of ionospheric imaging

Ionospheric tomography experiments have been performed since the early 1990s. In those days, the typical procedure was to reconstruct a two-dimensional image using measurements from a single satellite, with receivers arranged along the satellite track. The Navy Ionospheric Monitoring System (NIMS) satellites are in geographical polar orbit, so the receivers are placed close to lines of geographic longitude, whereas receivers for the Cicada satellites are placed in the equivalent geomagnetic locations.

Andreeva [1990] presented the first tomographic images of the ionosphere. The author used relative TEC observations from three locations in an iterative algorithm to reconstruct a two-dimensional image of the ionosphere. *Pryse and Kersley* [1992] also produced two-dimensional ionospheric images from three receiver sites and validated them with observations from the European Incoherent Scatter Radar (EISCAT). *Kersley et al.* [1993] showed that ionosonde observations could be used to set basis functions for the inversion algorithm and thereby improve the vertical profile agreement in comparison with EISCAT observations. *Kunitsyn et al.* [1994] compared the results of tomography using differenced phase observations with

results from calibrated observations. The authors found that biases were especially hard to calibrate in the presence of strong horizontal plasma density gradients and that differenced observations were more suitable for imaging localized electron density differences.

A joint Russian-American tomography experiment, held in 1993 and described by *Foster et al.* [1994], provided observations from both the NNSS and Cicada satellites that were used in two separate inversion algorithms. The American group followed a method described by *Raymund et al.* [1994], which uses a parameterized ionospheric model by *Daniell* [1991]. The imaging techniques successfully reproduced a depletion of ionization between 44°N and 66°N. *Markkanen et al.* [1995] used observations from Cicada satellites in a Bayesian approach. They produced two tomographic images, one of which showed similarities to observations from EISCAT, but they could not resolve the ambiguity in the vertical profile, so the issue of peak height determination remained unresolved. *Mitchell et al.* [1995] produced tomographic images from a chain of five receivers that showed irregularities aligned with the magnetic field. The images were produced using a MART algorithm with IRI-90 providing a background estimate of the ionosphere. *Bernhardt et al.* [1998] provides a review of ionospheric imaging using observations from the NIMS satellites.

3.7 Multi-Instrument Data Analysis Software

The ionospheric imaging approach used in Chapters 4, 5 and 6 of this thesis is a version of the Multi-Instrument Data Analysis Software (MIDAS). This section provides a detailed review of MIDAS, with particular attention given to the version used in this thesis.

3.7.1 Historic versions of MIDAS

Mitchell and Spencer [2003] created the first version of MIDAS. That version balances a series of spherical harmonics and empirical orthonormal functions derived from Chapman profiles around the world and assumes the ionosphere varies linearly with time. *Spencer and Mitchell* [2007] developed a second version using a Kalman filter

approach. This approach incorporates physics-based models in the inversion, meaning that it is a data assimilation approach as opposed to the earlier tomographic imaging method described by *Mitchell and Spencer* [2003]. The second version of MIDAS is particularly suited to imaging small-scale, fast-moving structures in the polar regions because the inclusion of a physics-based model allows the algorithm to overcome the historic lack of GPS observations near the poles. The lack of GPS observations in the polar regions has been dealt with in recent years, as was noted by *Kinrade et al.* [2013].

3.7.2 Introduction to the current version of MIDAS

The third version of MIDAS is used in this thesis. From this point onwards, "MIDAS" will be used to refer to this third version. MIDAS does not rely on a background model in the data assimilation sense – no weight is given to a modelled estimate of the state in the inversion, so this version should be considered a tomographic imaging approach. MIDAS solves for three-dimensional electron density distributions at multiple times from differential phase measurements of slant TEC. The algorithm relies on a vertical basis function decomposition and a horizontal smoothing function to make up for a lack of observations. Observations of differential slant TEC, derived from GPS phase measurements as described in Section 3.2.1, are normally used in MIDAS because they do not depend on the separate estimation of hardware biases.

In common with any tomographic imaging algorithm, the goal of MIDAS is to invert Equation 27, where A is the observation operator (in this case A is a matrix of GPS ray-path intersections with grid voxels), $\boldsymbol{x}$ is a vector of electron densities across the grid and $\boldsymbol{b}$ is a vector of observations.

$$A\,\boldsymbol{x} = \boldsymbol{b} \tag{27}$$

The goal is to invert Equation 27 in order to solve for the electron density distribution, $\boldsymbol{x}$, as is shown in Equation 28:

$$\boldsymbol{x} = A^{-1}\,\boldsymbol{b} \tag{28}$$

In a real problem, the number of independent observations is unlikely to match the number of grid-points – the problem will most likely be under-determined, meaning that there will be a lower number of observations than grid voxels (unknowns). The result of this underdeterminacy is that the observation operator, A, will not be a square matrix and so it will have no inverse. Equation 28 cannot be solved if the inverse of the observation operator, A^{-1}, is not known, so Equation 27 is left-multiplied by the transpose of the observation operator, A^{T}, and rearranged to give:

$$\boldsymbol{x} = (A^{T} A)^{-1} A^{T} \boldsymbol{b} \tag{29}$$

Equation 29 can always be solved, whether the problem is under-determined, over-determined or exactly determined. However, it can be difficult to solve this equation directly, so a least-squares approach is taken to minimise the residual:

$$\| A \boldsymbol{x} - \boldsymbol{b} \|^{2} \tag{30}$$

In the case of MIDAS, the problem is defined such that the grid is in three spatial dimensions and at multiple discrete times so that observations that span multiple times (differential phase observations of slant TEC) can be accommodated. However, Equation 29 can yield multiple solutions in cases where there are insufficient observations. Two different forms of mathematical constraint are applied to the inversion in order to restrict the possible solutions of the inversion to be physically realistic.

3.7.3 Strong constraints in MIDAS

A strong constraint is one that prevents any solution being returned that is outside of the favoured set. In the case of MIDAS, a mapping function is applied such that each vertical column of electron density is the sum of several (generally 2-4) orthogonal basis functions. These functions are designed to reproduce a series of modelled ionospheric vertical profiles – produced either from Chapman's equations or from the IRI-95 empirical model. The problem is simply mapped from grid space to basis function space, solved and then mapped back to the grid space for plotting and analysis. This approach has two main limitations: the basis functions do not force the solution to be positive, which densities of course should be, and they

prevent the algorithm from reproducing unusual ionospheric structures that are not present in the original models. The problem of negative electron densities is negated by good observation coverage, but can be avoided entirely by enforcing a non-negativity constraint in a computationally intensive non-linear optimization approach. Such an approach is an optional feature within MIDAS, but it is seldom used because it is so time-consuming. The problem of poorly selected basis functions can be avoided in case study situations by careful and knowledgeable users, but is dealt with automatically using ionosonde observations in Chapter 6.

3.7.4 Weak constraints in MIDAS

Weak constraints differ from strong constraints in that the solution can depart from a weak constraint if it means better agreement with the observations. Within MIDAS, a weak constraint is applied in the horizontal directions (latitudinally and longitudinally) and across the different times in the solution. The constraint chosen in MIDAS is a form of the well-known Tikhonov regularisation. This regularisation condition is designed so that solutions with zero second derivatives in these directions are favoured. The overall effect is to promote smooth solutions in regions lacking observations. The benefits and drawbacks of regularisation are explored in Chapter 5, using a simulated truth experiment. Regularisation allows MIDAS to produce accurate images in data-sparse regions, so long as the ionosphere is smooth, but can cause artefacts when the ionosphere is not smooth.

3.7.5 Use of models in MIDAS

MIDAS does not use a background model estimate of the solution in the data assimilation sense, but the algorithm is capable of using a modelled starting estimate of the solution to aid the minimization process. The idea is to begin optimizing from a starting guess that is likely to be close to the true solution in order to improve performance. The IRI-95 model is frequently used to provide a starting estimate at the beginning of a series of MIDAS images, after which the solution for the previous time-step is used. The IRI model is excluded from all aspects of MIDAS operation in Chapter 4 of this thesis, where MIDAS images are compared with a later version of IRI.

3.8 Summary

Ionospheric imaging techniques have developed from two-dimensional algorithms based on ray-path observations from fast-moving low-earth orbit satellites into three-dimensional, time-dependent methods that combine ground- and space-based GNSS observations with point measurements of electron density. Models are used to fill data gaps in data assimilation approaches, so model biases could lead to errors in those techniques. One widely used model, IRI 2007, is assessed in a long-term study that is presented in Chapter 4. In addition to providing direct estimates of ionospheric electron densities, models are used to create basis functions and regularisation conditions that restrict the solutions an algorithm can produce. These techniques allow algorithms to select only the realistic solutions in under-defined problems.

4 A comparison of tomographic and empirical ionospheric TEC specification

Abstract

In this chapter, the accuracy of current, state-of-the-art ionospheric specification techniques is assessed. The purpose of this study is to look for potential large, long-term biases in IRI 2007 using GPS-derived tomographic images of electron density over Europe and North America. The approach is to compare monthly median TEC estimates from an empirical model and from ionospheric tomography, produced hourly over the period 1998-2010 (one solar cycle). The results are validated with independent GPS slant TEC observations. Both techniques generally agree closely, but there are some discrepancies in IRI 2007 at solar maximum. These discrepancies are shown to be caused by unrepresentative coefficients of monthly median peak electron density, *NmF2,* used in IRI 2007.

The results of this chapter were published in *Chartier et al.* [2012a].

4.1 Introduction

Although large areas of the earth are now covered with dual-frequency GPS measurements, there is still not sufficient coverage to specify the ionosphere globally to a resolution of even a few hundred kilometres. This is one reason that ionospheric models are still important. IRI 2007 is an empirical model that specifies a global ionosphere. A combination of IRI 2007 and observations could be used to constrain the electron density within a coupled physical model. This would help overcome the problem of large data gaps over the oceans. However, the possibility exists that IRI 2007 could contain systematic biases. The comparisons shown in this chapter are performed using twelve years of data, from 1998 to 2010, in order to capture an entire solar cycle. A global study with lower resolution is conducted between 2005

and 2010, since data coverage is insufficient before then. Discrepancies between the MIDAS and IRI results are investigated using independent GPS slant TEC observations.

4.1.1 Ionospheric imaging techniques

Ionospheric imaging/data assimilation techniques provide an opportunity to mitigate the impact of active ionospheric conditions by providing delay corrections in near real time. Ionospheric imaging techniques are described in detail in Section 3.5. All these techniques depend largely on observations of slant TEC from GPS receivers. However, GPS data cannot cover the entire Earth and so there is a need to fill in information across data-sparse regions. The research described here evaluates the performance of IRI 2007 in terms of hour-to-hour monthly averages of vertical TEC. The evaluation is done using the MIDAS ionospheric imaging technique, because ionospheric imaging provides a consistent framework to calculate TEC maps without the distortions created by mapping functions [*Meggs et al.*, 2005]. MIDAS is described in detail in Section 3.7.

4.1.2 The International Reference Ionosphere

IRI 2007 is an empirical model of the ionosphere based on a wide range of ground and space data, including incoherent scatter radars and topside sounders. It is the result of collaboration between the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI) that began in 1969. It produces monthly median reconstructions of electron density, ionized gas composition and temperature in the altitude range 50 – 1500 km (65-2000 km for electron density). The 2007 model has new topside electron density options designed to improve TEC measurements by removing an artefact that overestimated electron densities above 500 km. IRI 2007a subtracts a correction coefficient from the IRI 2001 values above 500 km while IRI 2007b uses a different topside model, also used in NeQuick. This model was developed by *Radicella and Leitinger* [2001]. IRI 2007b produces better results than IRI 2007a according to four tests run by *Bilitza and Reinisch* [2008].

4.2 Method and Results

4.2.1 12 year comparison of MIDAS and IRI 2007

Two twelve year IRI runs were performed, one with the NeQuick topside model switched on (IRI 2007b) and another with it switched off (IRI 2007a). There was minimal difference (around 1 TECU, or 10^{16} electrons/m^2) between the two at any time in terms of a bulk TEC number across the entire grid. The NeQuick (IRI 2007b) runs were chosen for the following comparisons because *Bilitza and Reinisch* [2008] showed that it performed better. The top of the grid was set to the maximum 2000 km to minimise the plasmaspheric difference when comparing with GPS data (GPS satellites are at 20,100 km). IRI aims to produce a monthly average of the ionosphere for magnetically quiet conditions, so thirty-day median hourly MIDAS images were created to give a fair comparison with IRI. For example, an IRI image for 12:00 Universal Time (UT) on the 15th of April was compared with the cell-by-cell thirty-day median of the MIDAS vertical TEC images at 12:00 UT in April. The mean TEC across the cells of this median image was compared with the mean TEC in the IRI image to check for systematic biases, while the RMS difference between the two was calculated to assess the spatial variability.

Regions of good GPS data coverage were identified in North America and Europe between 1998 and 2010. At least 40 sites were available in each of these areas for most of that period. This allowed the observation of the ionosphere over a full solar cycle. GPS receiver site lists for North America and Europe were updated every three years to deal with new sites being installed and old sites being switched off. This meant the spatial distribution of the data improved over the 12 years. The exact number of sites in use varied due to temporary outages. The impact of this was difficult to assess, because closely grouped sites were less useful to the reconstructions than ones that were more spread out. Reconstructions were performed hourly to capture the major diurnal variations. GPS data were supplied by the International GNSS Service (IGS). See *Dow et al.* [2009] for more details.

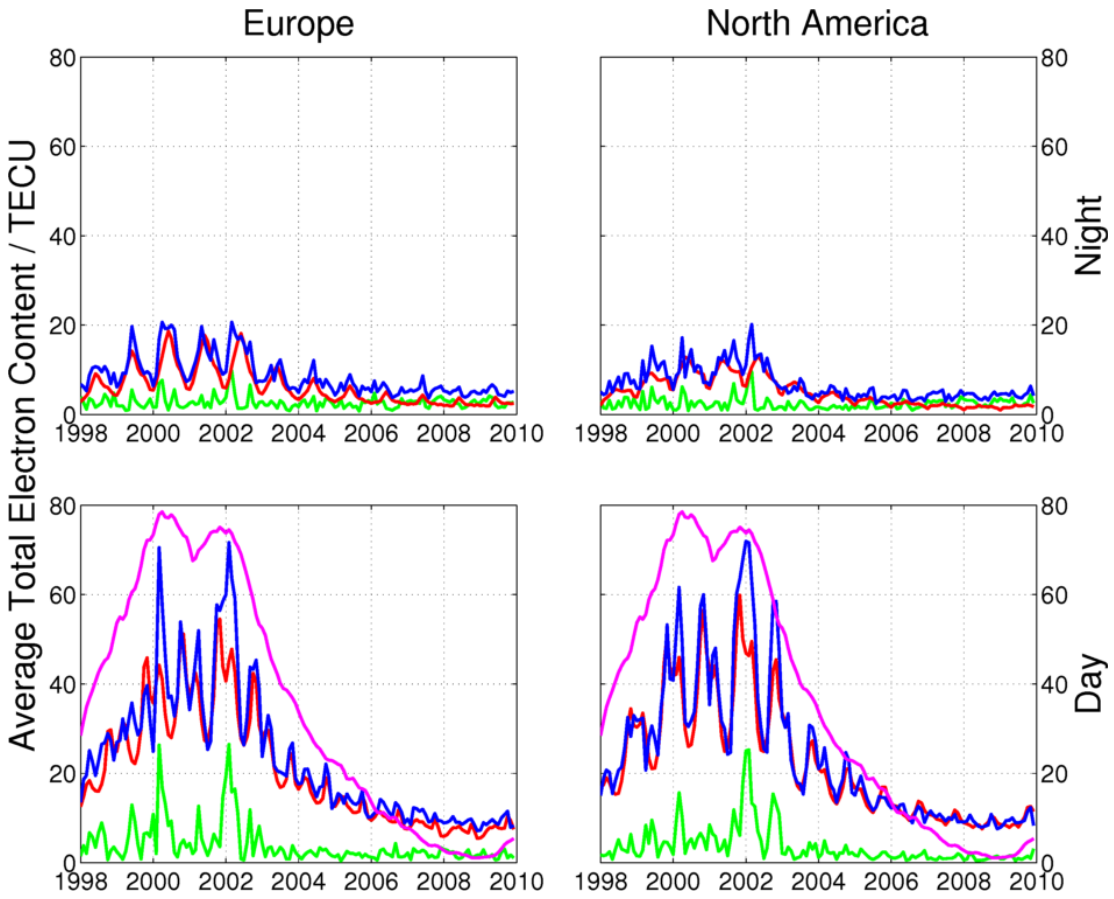

Figure 5: The mean TEC of IRI (red) and 30-day median MIDAS (blue) images between 1998 and 2010 over Europe and North America are shown. Day-time images are based around local noon, whilst night-time images are based around local midnight. The green line shows the RMS difference between the images. The purple line in the day-time images shows the smoothed sunspot number reduced by a factor of 1.5.

Figure 5 shows close agreement between MIDAS and IRI over most of the 12-year period in Europe and North America. Day-side disagreement peaked in spring 2000 and particularly spring 2002, with MIDAS values significantly higher. Night-side plots showed no such peaks in disagreement, but did show slightly higher TEC for much of the 12 year period. Root-Mean-Square (RMS) differences were low most of the time, showing that TEC maps were similar across the grid. Peaks in RMS difference were associated with peaks in overall difference between the two images. Spring 2000 and spring 2002 peaks in disagreement coincided with peaks in sunspot number. This may be an indication that the 10.7 cm solar flux, *F10.7*, used by IRI did not sufficiently characterise solar irradiance. However, the autumn 2002 peak in disagreement in North America did not fit this trend. After 2003, MIDAS and IRI TEC generally agreed well in the regional plots.

4.2.2 Systematic Bias

Large differences between MIDAS and IRI are evident in the daytime regional reconstructions around solar maximum. Combined GPS satellite and receiver bias estimates were extracted from the MIDAS images and compared with estimates from the Centre for Orbit Detection in Europe (CODE). Images from February 2002 were tested since this was the month where MIDAS showed the greatest difference with IRI. The results are shown in Figure 6. Images from April-September 2004 were also tested as a control. These results are shown in Figure 7. IRI images could not be directly tested, since IRI aims to provide only a monthly median of conditions.

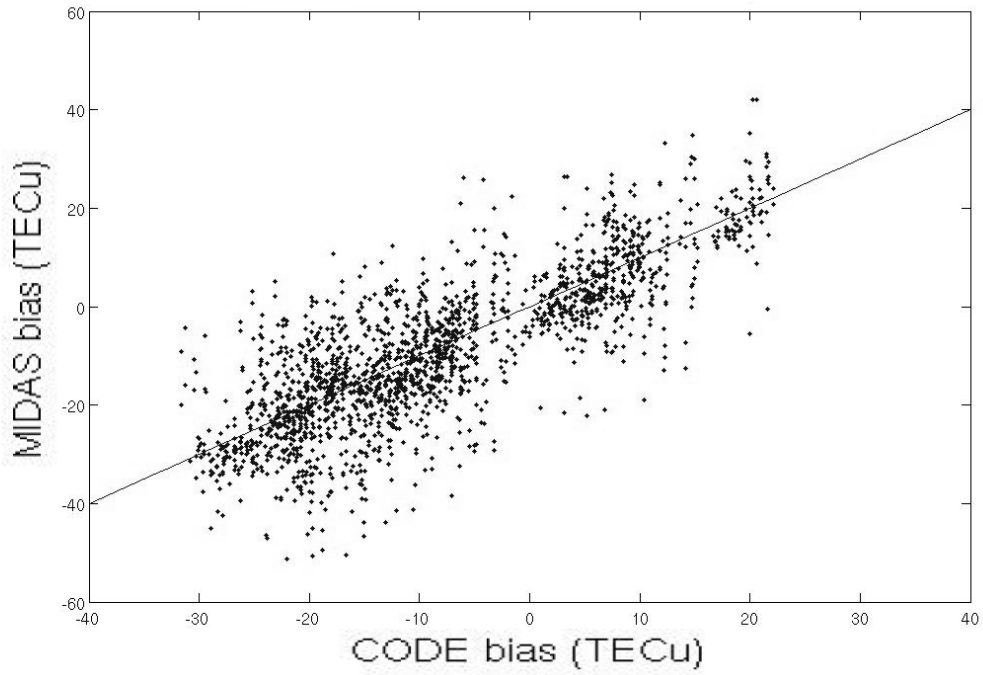

Figure 6: MIDAS and CODE estimates of combined satellite-receiver biases in February 2002, a period of intense geomagnetic activity, are compared.

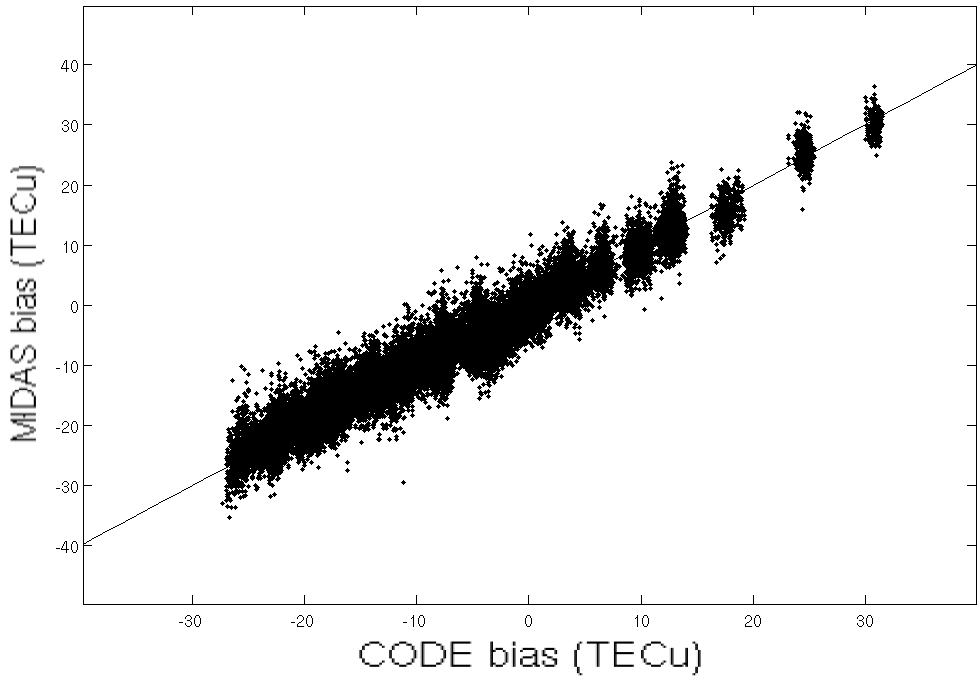

Figure 7: MIDAS and CODE estimates of combined satellite-receiver biases in April-September 2004, a period of moderate geomagnetic activity, are compared.

The technique for deriving bias estimates from MIDAS involved subtracting independent estimates of slant TEC from the MIDAS slant TEC values. The independent data were taken from GPS phase-smoothed pseudorange estimates of TEC. These measurements provide absolute TEC values, but contain significant hardware biases from both the satellites and the receivers (see *Choi et al.* [2011] and *Wilson and Mannucci* [1993]). The difference between the MIDAS and GPS slant TEC values will be the hardware bias, so long as there is no remaining bias in the MIDAS image.

Singular value decomposition was used to solve for the individual satellite and receiver biases across the network. These estimates were then added back together to form combined satellite-receiver biases. This last step was necessary because there was no way of determining how the biases were shared between satellite and receiver.

The MIDAS bias estimates were compared with estimates from CODE. *Hugentobler et al.* [2000] state that CODE produce inter-frequency biases for all GPS satellites and

around 160 IGS receiver stations that are repeatable to around 0.285 TECU (converted from nanoseconds, as per *Gaposchkin and Coster* [1993]). This test shown in Figure 6 indicated that the MIDAS images for Europe in February 2002 had no systematic bias in TEC. Similar results were observed in the North American reconstructions at this time, although the number of receivers for which CODE bias estimates were available was limited in that region. The results in Figure 7 showed that no systematic bias was observed in April-September 2004 either. There was less random error in these measurements because the ionosphere was less variable in that period. These results indicate that the MIDAS results were not biased.

The bias in IRI TEC was investigated by examining monthly median TEC over the Rome ionosonde station in March 2002. IRI was run using manually scaled ionosonde observations of peak height (*hmF2*) and peak plasma frequency (*foF2*) and, separately, using the default coefficients. The default mode, as used in the other tests, underestimated the ionosonde-corrected TEC by 52 TECU. *foF2* inaccuracies were by far the most important source of error in this case. This demonstrates that the large systematic errors observed in IRI can be caused by poor peak data from the coefficients. Ionosonde observations were provided by the Electronic Space Weather and Upper Atmosphere (eSWua) database of the Istituto Nationale di Geofisica e Vulcanologia (INGV), as described by *Romano et al.* [2008].

4.2.3 Diurnal Variation

Figure 8 shows the mean diurnal TEC variation in Europe over the 12 year period. Both datasets showed the same overall pattern, with MIDAS consistently around 2 TECU higher than IRI. Similar results were observed in North America. These results were to be expected since the IRI simulations span 0 – 2000 km, whilst the GPS rays used as input to MIDAS contain information on the plasmasphere. It is clear that this plasmaspheric content in GPS TEC observations cannot be ignored in a data assimilation scheme. *Coster et al.* [1990] observed over 30 TECU between 800 km and the GPS satellites (20,100 km) during a period of variable geomagnetic activity near solar maximum. *Ciraolo and Spalla* [1997] found an overall value of 3 TECU between 1100 km and 20,000 km by comparing GPS and NNSS TEC measurements over two years.

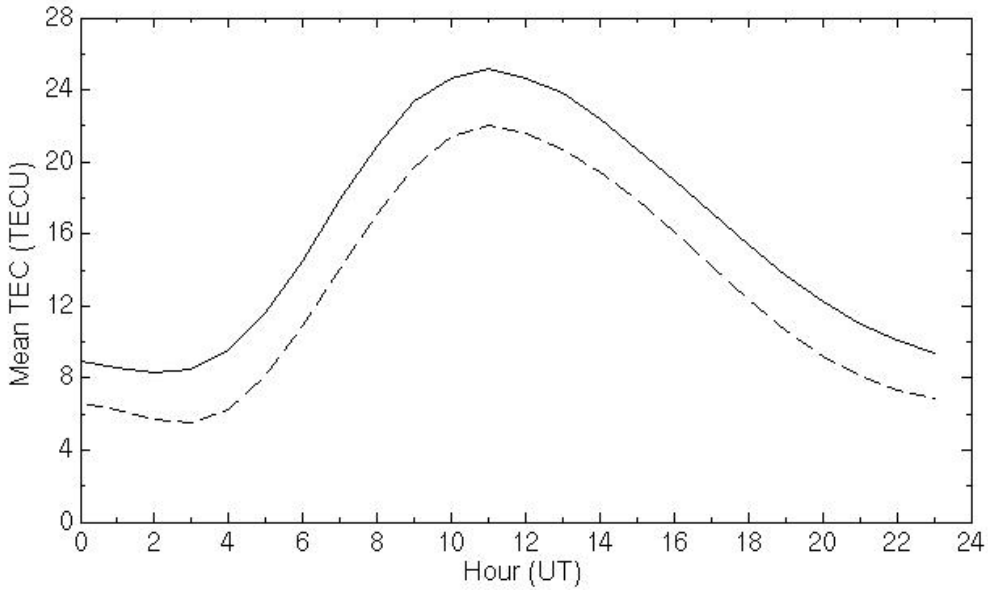

Figure 8: The mean diurnal variation of TEC, based on European data from 1998 to 2010. The dashed line indicates IRI whilst the solid line indicates MIDAS.

4.2.4 Global reconstructions

Near-global-scale reconstructions were performed for MIDAS after 2005. Prior to this date, there were insufficient data to perform a numerically stable global-scale inversion. Reconstructions were performed hourly, using 100 sites and with 5x10 degree horizontal resolution. Polar caps (>80 degrees North or South) were omitted from the global reconstruction because IRI 2007 does not perform well in the polar region. The MIDAS images used in this test were limited by a lack of GPS data over the oceans. This was mitigated by the use of lower than normal horizontal resolution and a longer time-window, but those steps meant the inversion coped less well with the variable conditions seen before 2005. Figure 9 shows the mean TEC of the IRI and monthly median MIDAS images from January 2005 to December 2009.

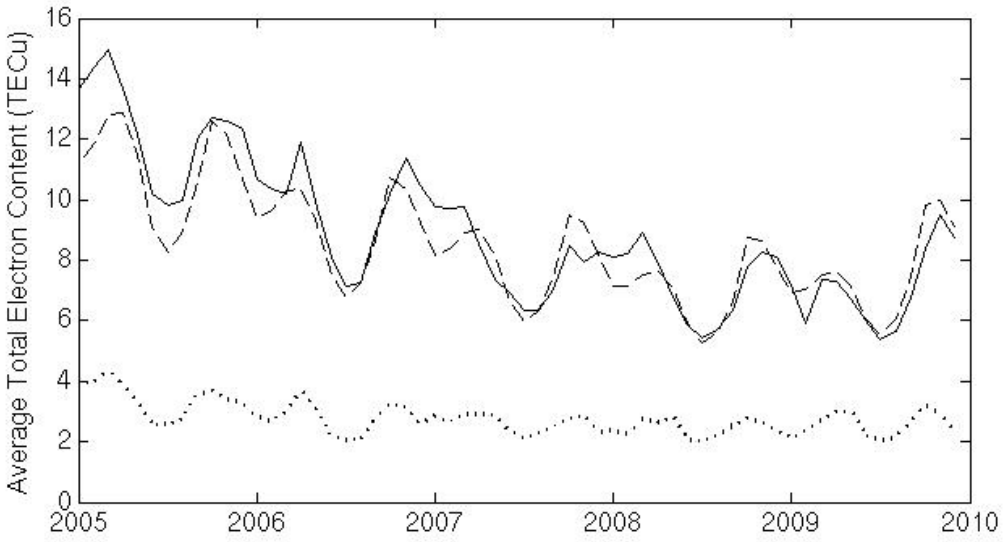

Figure 9: The mean TEC of IRI (dashed line) and 30-day median MIDAS (solid line) images between 2005 and 2010 over all longitudes and up to 80° North and South. The dotted line shows the RMS difference between the images.

Various sources (e.g. *Rishbeth and Muller-Wodarg* [2006]) report an annual asymmetry in *NmF2* values. A 30% difference in *NmF2* is observed between January and July, much more than could be explained by the 7% annual variation in solar flux which is due to orbital variations. It appears there is more plasma in January than in July. *Mendillo et al.* [2005] show that the asymmetry exists in TEC as well as in peak density. They exclude the possibility that plasma is simply being transported above the peak height - and therefore out of sight of the ionosondes - by comparing TEC from GPS data in December 2001 and June 2002 between 65 degrees north and 65 degrees south. *Mendillo et al.* [2005] obtain an asymmetry index, *AI*, of 0.15 using the following calculation:

$$AI = [Dec01 - Jun02] / [Dec01 + Jun02] \quad (31)$$

This is a reasonable index to use over periods greater than one solar cycle. However, the index is susceptible to changes in solar activity over periods of a few years or less because of solar cycle behaviour. For example, December 2004 should have less global TEC than June 2003 since sunspot numbers decreased over the same period. The following index, shown in Equation 32, removes the effect of solar cycle behaviour. For consistency with the earlier ionosonde measurements of asymmetry, January and July are used here instead of December and June. Each year's

asymmetry index is computed from that January's data, and from the mean of the preceding and following July. In the case of 2002, we calculate the asymmetry index, *AI*, as follows:

$$AI = [\ Jan02 - mean\ (\ Jul01 + Jul02\ )\ ]\ /\ [\ Jan02 + mean\ (\ Jul01 + Jul02\ )\ ] \qquad (32)$$

This method was used by *Rishbeth and Muller-Wodarg* [2006] to remove the solar cycle behaviour. MIDAS results for TEC asymmetry were quite similar to *Rishbeth and Muller-Wodarg's* [2006] *NmF2* asymmetry from pairs of ionosondes. Figure 10 shows that the MIDAS asymmetry index varied between 0.08 and 0.15 between 2006 and 2009, which is generally somewhat lower than *Mendillo et al's* [2005] result of 0.15. This is most likely because we removed the solar cycle behaviour from our results. IRI shows similar behaviour between 1999 and 2009, with an index of between 0.08 and 0.14. IRI's asymmetry index, which represents the asymmetric content as a proportion of total ionization, remains roughly constant over the solar cycle. This means the asymmetric effect has varied at the same rate as the total plasma content.

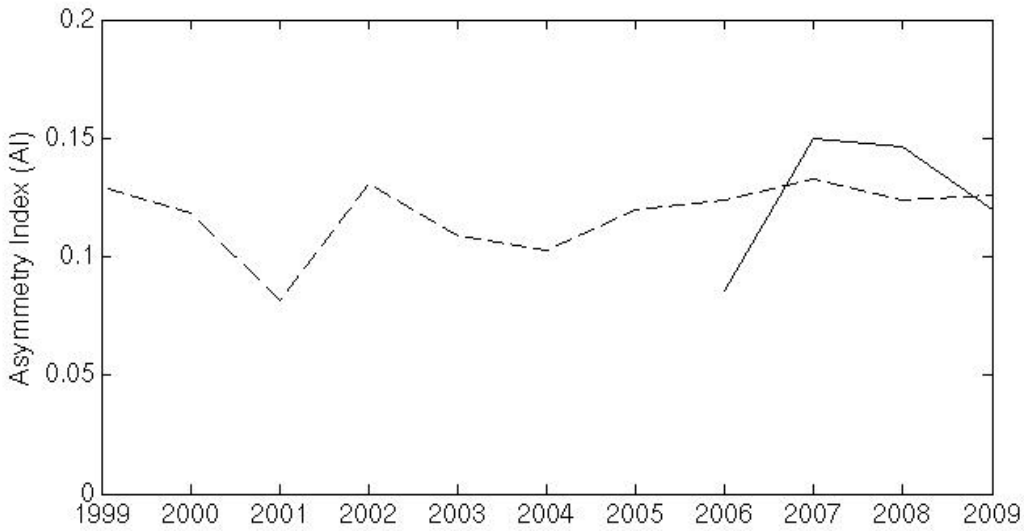

Figure 10: The annual asymmetry index (AI) taken from global IRI 2007 (dashed line) and 30-day median MIDAS (solid line) images in January and July. The index gives an indication of the imbalance in TEC over this period. An index of zero would indicate no imbalance in TEC.

4.3 Discussion

A twelve year comparison of IRI 2007 with monthly median GPS derived tomographic images of the ionosphere showed that the empirical model closely matched TEC over Europe and North America for most of that period. This demonstrated that monthly median ionospheric TEC at mid-latitudes can be specified by the limited range of global geomagnetic and solar parameters used by IRI 2007. Therefore it is feasible that GPS data could be combined with IRI 2007 to define a global ionospheric analysis with variable structures in regions of good data coverage. It would be beneficial for such an analysis to be used to constrain the ionospheric element of a coupled physical model of the upper atmosphere and feed that information to the little-observed thermospheric state. This in turn would allow the physical model to produce improved forecasts of both charged and neutral parameters. Forecasting the upper atmosphere is an important challenge, since it will allow us to improve the resilience of GNSS, reduce risks to flights over the polar caps and help to more accurately determine satellite orbits.

The comparison showed that IRI 2007 significantly underestimated TEC in February-March 2000 and 2002, compared to MIDAS images validated with independent GPS data. This was most likely caused by URSI and Comité Consultatif International pour la Radio (CCIR) coefficients underestimating *hmF2* and *NmF2* at these times, since the use of ionosonde-derived *hmF2* and *NmF2* increased IRI TEC over Rome by 52 TECU in March 2002. The authors look forward to the development of a data assimilation scheme for IRI that could incorporate ionosonde data into the three-dimensional electron density images. In the meantime, the assessment of IRI peak errors can be used as an indication of errors in TEC.

IRI's underestimation of TEC by about 2 TECU over the whole 12-year period may be explained by topside plasma, but it has not been possible to determine whether this is above or below the 2000 km top of the model. The sources referenced in Section 4.2.3 show that there could be at least 2 TECU above 2000 km, the top of IRI. The alternative scenario is that the extra plasma is contained within the limits of IRI and that therefore the model is in error. If that is the case, the bias is likely to be in

the topside, since IRI is largely based on observations of the bottom-side and peak. This is supported by *Lee and Reinisch* [2006], who found that IRI 2001, an earlier version of the model, had no significant *NmF2* biases when compared to ionosondes at Jicamarca. In addition, *Jakowski and Mayer* [2009] found that IRI 2007 with NeQuick underestimates topside electron densities at low and mid latitudes from 400 km up to 2000 km. The data from this study came from assimilative reconstructions of CHAMP satellite GPS data between 2002 and 2006, which also indicated plasmaspheric TEC. In order to develop the IRI project, it would be beneficial for work to be done on improving our understanding of the topside plasma distribution, perhaps using TEC measurements from a satellite at an altitude of 2000 km or thereabouts. If significant quantities of plasma are found above that height, the model top will need to be extended, whereas a finding of only negligible densities would indicate a bias in underestimation of densities below that height.

IRI was often seen to be biased in the equinoctial months, both in the regional and the global reconstructions. This bias was not always in underestimation - Figure 10 shows an overestimation around September 2007 followed by an underestimation leading up to March 2008. Global RMS errors also consistently increased around the equinoxes. It is possible that IRI failed to represent the more active conditions in this period, since the model was primarily designed for use in geomagnetically quiet conditions [*Bilitza and Reinisch*, 2008]. *Kunches and Klobuchar* [2001] observed more magnetic storms during the equinoctial months because of the relative orientation of Earth's magnetic field within the interplanetary magnetic field. IRI's error may also be due to an inability to represent some of the processes underlying the semi-annual asymmetry. This phenomenon, observed by *Pham Thi Thu et al.* [2011] and others, is distinct from the annual asymmetry referred to earlier and instead describes enhanced ionization in the equinoctial months. If the semi-annual asymmetry is caused by composition changes in the neutral atmosphere, then it could be the cause of the biases in IRI at these times. This is because the topside model, NeQuick, contains no information on neutral atmospheric composition, as explained by *Coisson et al.* [2005]. This explanation requires that the neutral atmospheric composition plays an important role in the topside ionosphere, specifically around the equinoxes. Possible causes, then, include *Rishbeth and Setty's* [1961] explanation, a change in the O/N2 ratio, and *Zou et al.'s* [2000] suggestion that the asymmetry

could be due to tides propagating up from the lower thermosphere. These are not necessarily the main causes of the semi-annual asymmetry, but their variability could be the reason for the errors we observed in IRI. *Jee et al.* [2005] observe that IRI 2001, an earlier version of the model, had much less semi-annual asymmetry than TOPEX ocean altimetry TEC data during periods of high solar activity. A consideration of thermospheric parameters may help IRI cope with the ionospheric variability caused by these processes.

In summary, monthly median estimates of TEC from MIDAS and IRI-2007 in Europe and North America were shown agree well for most of the twelve year period of study. The MIDAS images were validated and shown to be correct when the two approaches differed, with errors in IRI-2007 due to inaccurate input coefficients. The following chapters will explore different methods of improving ionospheric specification techniques, starting with the problem of imaging horizontal TEC structures in regions of sparse observation coverage.

5 Ionospheric imaging in Africa

Abstract

In this chapter, the problem of imaging ionospheric TEC in regions of poor observation coverage is addressed. A simulation approach demonstrates that current operational GPS receiver coverage is insufficient to reproduce the horizontal structures shown in a model of TEC over Africa. Image quality is substantially improved by the inclusion of observations from an extended receiver network. A fictional, high-density receiver network produces some further improvements to image quality, but some errors remain. This shows that observations from other sources than ground-based GPS receivers are required if imaging errors are to be eliminated.

The results of this chapter were published in *Chartier et al.* [2013a].

5.1 Introduction

Human activities such as high-frequency communications and satellite positioning are affected by ionospheric electron densities. Ionospheric imaging could be used to estimate effects on these activities in the African sector. The work presented here utilizes a simulation approach to determine the accuracy of tomographic images of the ionosphere. This approach allows application developers to quantify the benefits of including observations from sites that are not currently operational. Real images are also presented and analysed.

5.1.1 Ionospheric tomography

GPS ionospheric tomography techniques invert observations of relative or calibrated slant TEC from dual-frequency GPS receivers to produce three-dimensional, time-dependent images of electron density. The resulting ionospheric images can be used to correct ionospheric errors in GPS positioning - *Allain and Mitchell* [2009] showed

that a real-time version of MIDAS could improve single-frequency GPS position estimates by up to 25 m at mid-latitudes during solar maximum. The version of MIDAS used here is described in detail in Section 3.7.

5.1.2 Ionospheric observations

In common with other ionospheric imaging techniques, MIDAS depends on good data coverage and utilizes certain assumptions to fill data gaps. Ionospheric imaging in the African sector presents a special challenge since there is a large gap in GPS receiver coverage in the Sahara (see Figure 11). The observations used in this study are provided by three networks: IGS, the University Navstar Corporation (UNAVCO) and the AFrican geodetic REference Frame (AFREF). The IGS network is described by *Dow et al.* [2009].

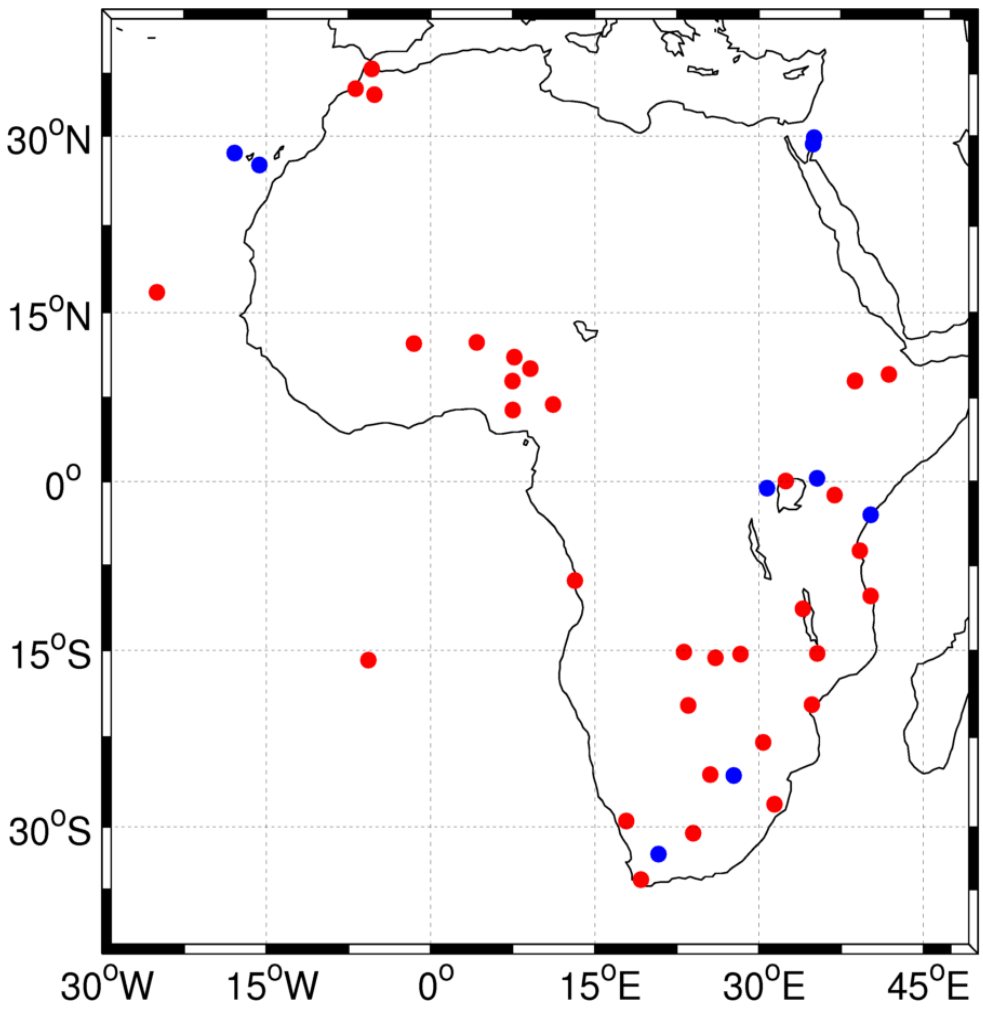

Figure 11: Working IGS network dual-frequency GPS receivers in blue and additional sites from the AFREF and UNAVCO networks in red. Only sites that produced usable observations in the period 2-16 December 2012 are shown here.

Figure 11 shows that operationally available sites (those provided by the IGS and those from UNAVCO) cover northern and southern Africa quite well, but there is little data available between 15°N and 25°N. The AFREF project is primarily aimed at unifying geodetic reference frames for Africa, but also recognizes other applications of GNSS signals, such as for ionospheric imaging. By including data from these receivers in simulated inversions, it is possible to assess the impact of observations that come from geographically feasible locations. A comparison can be made between the quality of ionospheric reconstructions from existing data and the improvements that can arise from including additional receiver stations.

5.1.3 Ionospheric models

The work described in this chapter relies on realistic simulations of ionospheric electron density to serve as a ground truth. The latest version of IRI, IRI-2012, is chosen for this purpose. The observational noise that is to be expected from sub-grid scale structures is dealt with separately (see Section 5.2.3). IRI-2012 is an empirical model of the ionosphere based on a wide range of ground and space data, including incoherent scatter radars and topside sounders [*Bilitza et al.,* 2011]. It is the result of collaboration between COSPAR and URSI that began in 1969. IRI models the monthly median electron density, ionized gas composition and temperature in the altitude range 50 – 1500 km.

5.2 Method

The quality of ionospheric reconstructions is difficult to assess without an independent ground truth. In this experiment, we use a modelled ionosphere from IRI-2012 as a ground truth. Simulated GPS TEC observations through this modelled ionosphere are created and used in MIDAS inversions. As well as allowing for comparison of the images with the ground, this approach allows us to quantify the benefits of including observations from additional sites that are not currently operationally available. A similar approach was used by *Dear and Mitchell* [2006] in Europe and by *Zapfe et al.* [2007] in South America. Differences between the model and the reconstructed images are then due to a lack of observations, or poor assumptions about the nature of the solution. Whilst the MIDAS inversion technique

can use IRI to create basis functions and to provide an initial guess at the solution, those options are disabled for this experiment. Instead, we use two basis functions derived from Chapman's equations and a 4^{o} horizontal grid. In addition to assuming the ionosphere can be adequately represented by these constraints, a regularisation condition is applied that favours solutions with zero second derivative in each horizontal direction and in time. These assumptions are necessary to obtain a unique solution and to cover data gaps, but it is noted that the assumptions could prove problematic if they are not appropriate for a specific ionosphere. The selection of IRI-2012 as the 'truth' ionosphere might artificially enhance the performance of the imaging algorithm because IRI-2012 is smoother than the real ionosphere and our algorithm favours smooth solutions. This problem is addressed by adding realistic noise, caused by sub-grid level structures, to the observations. The addition of realistic noise to simulated observations is described in Section 5.2.3 and Section 5.2.4, but it should be noted that the problem of scintillation is not considered.

5.2.1 Simulating the ionosphere

IRI-2012 is used to create three-dimensional fields of electron density values on the same four-degree grid that will be used for the reconstructions. MIDAS uses observations from a time-window around the inversion, so it is necessary to simulate the ionosphere for multiple times. In this case, we use a time-window of seven hours and thirty minutes with a time-step of thirty minutes. The result is that, for each inversion, fifteen IRI simulations are created at thirty-minute intervals. The resulting IRI electron density values are arranged into a simulated state vector, $\boldsymbol{x}_{\mathrm{IRI}}$.

5.2.2 Simulating a receiver network

In order to find the upper limit of imaging accuracy possible using the chosen grid for the MIDAS algorithm, it is necessary to simulate a network of receivers that would provide adequate observation coverage. This is achieved by creating a regularly spaced list of coordinates to represent fictitious receivers at eight degree intervals in latitude and longitude. This receiver spacing was chosen because it was found that increasing receiver coverage above eight degree spacing made almost no difference to image accuracy – almost identical images produced from a four degree spaced network. This list of coordinates is combined with the known position of the

GPS satellites to find the trajectories of the rays that would be observed by the simulated receiver network. Real receiver networks can also be used in the simulation approach, with the added advantage that data outages can be taken account of. This is described in Section 5.2.3.

5.2.3 Simulating observations

As already noted, it is necessary to create TEC 'observations' of the simulated ionosphere in order to produce reconstructed images. This is achieved using an observation operator, *A*, that is based on the trajectories of real or simulated observations. The observation operator, *A,* describes the ray-path contributions of the observations to the grid. The operator *A* is created by tracing ray-paths from the GPS satellites to the receivers at the times when data is received. In order to create a vector of simulated observations, $\boldsymbol{b}_{\mathrm{IRI}}$, we multiply the simulated state vector by the observation operator:

$$\boldsymbol{b}_{\mathrm{IRI}} = A\,\boldsymbol{x}_{\mathrm{IRI}} \tag{33}$$

We assume that real GPS differential phase observations of slant TEC contain significant errors, although they do contain cycle slips. It is possible that the observations are not representative of grid-scale structures – there could be significant sub-grid-level 'noise' in the real ionosphere. This could happen if structures exist in the ionosphere that are too small to image using the specified resolution. Our simulated ionosphere, IRI-2012, will not have these sub-grid scale structures because it is defined on the same grid that will be used in the inversion. The simulated ionosphere is also far smoother than the real ionosphere. It is important to include realistic errors of representativeness in the simulation so that the inversion accuracy is not artificially enhanced. This is achieved by creating an image, $\boldsymbol{x}_{\mathrm{REAL}}$, of the real electron density distribution using real observations of slant TEC, $\boldsymbol{b}_{\mathrm{REAL}}$, and then calculating the residuals, $\boldsymbol{r}$, of the observations from the images:

$$\boldsymbol{r} = \boldsymbol{b}_{\mathrm{REAL}} - A\,\boldsymbol{x}_{\mathrm{REAL}} \tag{34}$$

These residuals are added to the simulated observations of slant TEC in order to take account of the effects of sub-grid level structures on image accuracy.

5.2.4 Reconstructions using simulated observations

The simulated observations are inverted in order to reconstruct the simulated ionosphere. The normal MIDAS inversion procedure is followed. First, the problem is mapped from electron density space to basis function space. A mapping function, M, is used to make the transformation. Then, a regularisation condition, R, is applied that penalizes solutions that contain non-zero second derivatives of electron density in horizontal space and in time. In practice a weighting term, w, has to be included in order to balance the effects of the measurements and the regularisation on the solution. Within MIDAS, a heuristic choice is made to define the regularisation weighting as:

$$w = trace\,(\,M^{T}\,A^{T}\,A\,M\,)\;/\;trace\,(\,R\,) \tag{35}$$

Finally, the constrained problem is inverted to obtain a solution, $\boldsymbol{x}_{\text{RETRIEVED}}$, for all the times in the time-window:

$$\boldsymbol{x}_{\text{RETRIEVED}} = (\,(\,M^{T}\,A^{T}\,A\,M\,) + w\,\mathrm{R}\,)^{-1}\,M^{T}\,A^{T}\,\boldsymbol{b}_{\text{IRI}} \tag{36}$$

The central slice in time is selected as the final image. The retrieved solution, $\boldsymbol{x}_{\text{RETRIEVED}}$, is compared with the original ionospheric simulation, $\boldsymbol{x}_{\text{IRI}}$, in order to determine the accuracy of the imaging technique given the available data.

A two-week period of the recent AFREF campaign provided a great deal of extra GPS coverage in the African sector (see Figure 11). This provides the opportunity to contrast the image quality possible using the existing IGS network with the image quality that an extended network could provide. The procedure outlined above is run twice: once using just the available IGS sites and a second time supplementing this with the sites available through UNAVCO and AFREF. Simulation inversions are performed for the maximum AFREF data availability period of 2 - 16 December 2012. Differences between the two sets of images show the improvements in accuracy that can be achieved by using additional receivers.

5.3 Results

Following the procedure described in Section 5.2, three sets of simulated inversions

were produced for the maximum AFREF data availability period (2 – 16 December 2012). The first set of images is based on the simulated, regularly spaced receiver network. These results provide an estimate of the upper limit of imaging accuracy achievable using the MIDAS algorithm on a 4° grid and under ideal conditions. The other two sets of images are created to determine the imaging accuracy achievable using real GPS receiver networks. One set of images was based on simulated observations at the locations of the IGS receivers, whilst the other set was based on all the available receivers. In each case, images were produced every 30 minutes throughout the test period.

5.3.1 Imaging under optimal conditions

Although insufficient observation coverage is likely to be the primary source of error in ionospheric images, it is possible that inherent properties of the imaging technique also limit accuracy. The results presented in this section demonstrate the performance of the imaging technique when provided with high density (8 degree spaced) and uniform GPS receiver coverage. 100 receivers are used in total. An example of the model truth and the image obtained from this high density simulated network is shown in Figure 12.

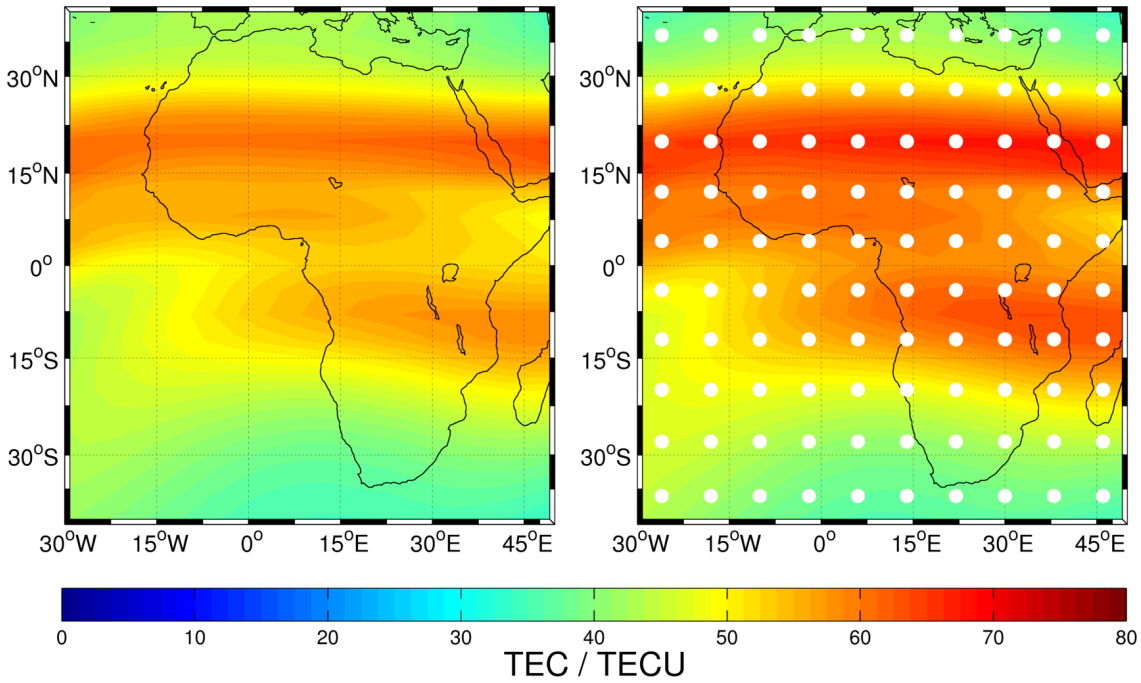

Figure 12: Left: IRI modelled truth. Right: reconstructed image based on observations of the modelled truth from a fictitious receiver network (shown in white). The model and reconstructed image are from 12:00 UT on 7 December 2012.

Images such as the one in Figure 12 are produced at 30-minute intervals over the period 2 – 16 December 2012. In order to assess the errors of the images, differences between the reconstructed images and the modelled truth are calculated over the whole period. The RMS errors of the images from the fictitious receiver network are shown in Figure 13.

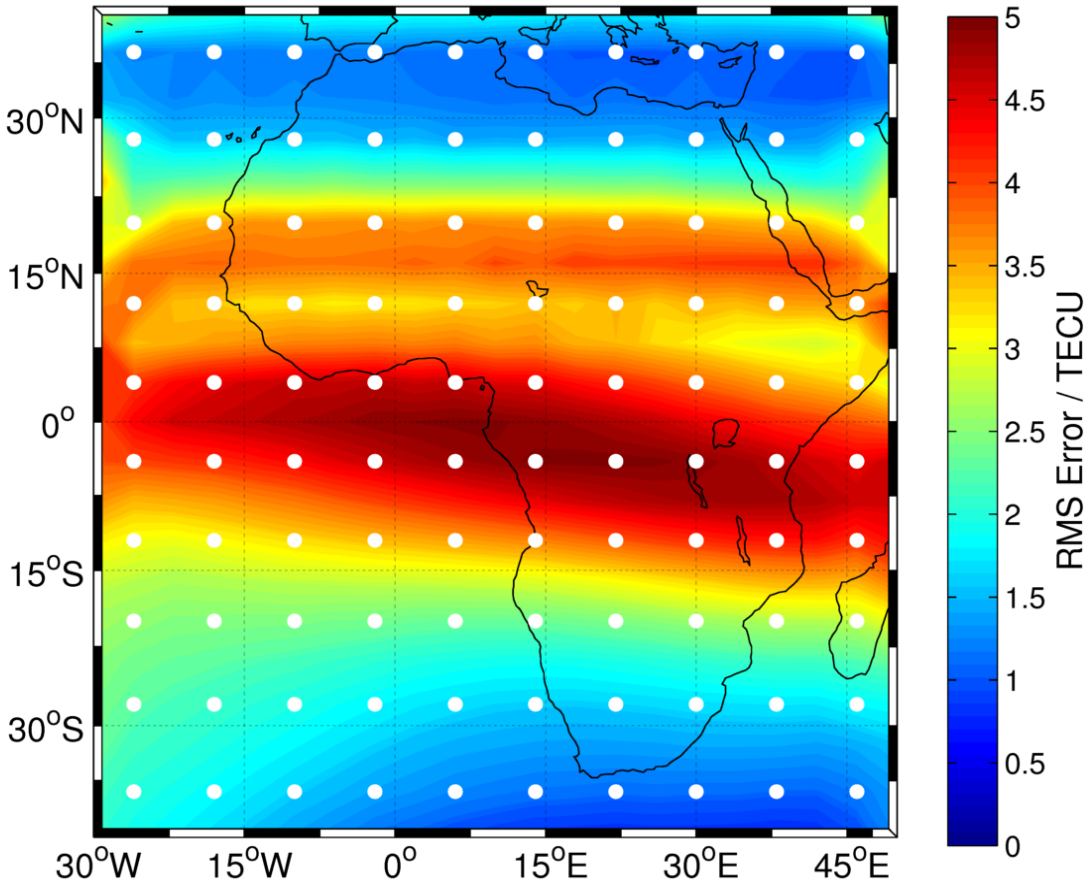

Figure 13: The RMS errors of reconstructions based on simulated observations from a dense network of simulated receivers are shown here. The simulated receiver sites are shown in white. The IRI simulations that the observations are based on are used as the ground truth here. The RMS errors are based on hourly reconstructions from the two-week test period (2-16 December 2012).

The results of Figure 13 show that relatively small errors can be achieved when a dense network of receivers is available – errors range from 0-5 TECU here. Errors are clearly highest at the locations of the two bands of increased ionization created by the Appleton anomaly (around 10°S – 0° and 15°N – 20°N). As well as the higher TEC values present here, the TEC gradients seen in this region are likely to

contribute to these higher errors. The imaging technique includes a regularization condition that favours zero second derivatives in the horizontal directions – a condition that is clearly broken as we move over the Appleton anomaly in a longitudinal direction.

5.3.2 Data coverage

In order to illustrate the typical data coverage available for each set of images, a pair of case studies are shown in Figure 14. The 300 km pierce points of all the rays available for use in each image are shown. These rays are collected at 30 minute intervals over a 7.5 hour time-window. The case study shows data from 0:00 – 7:30 UT on 3 December 2012.

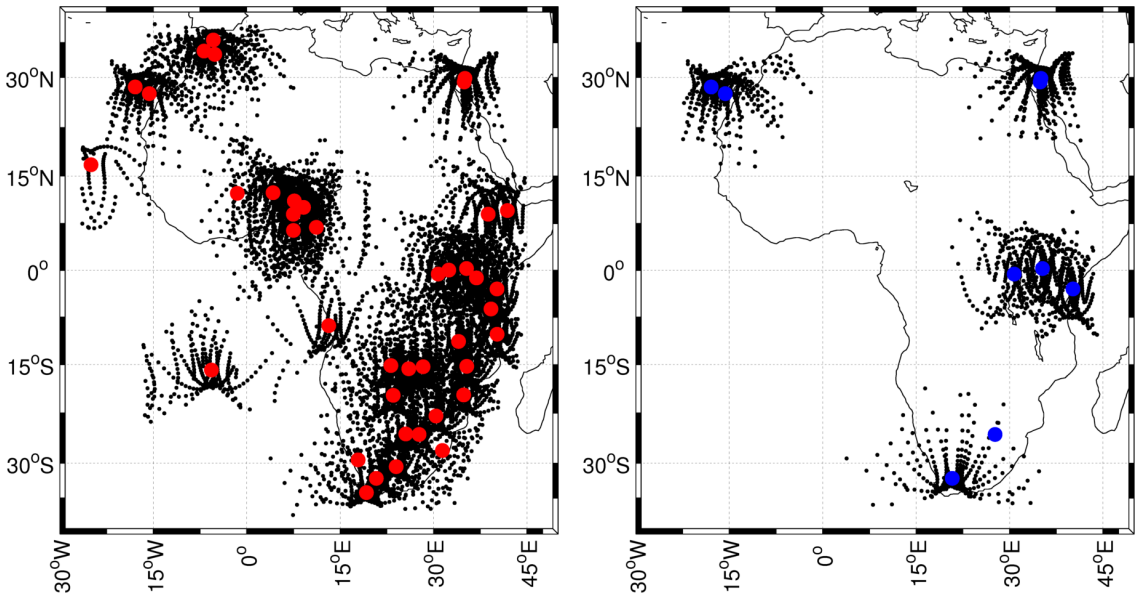

Figure 14: The GPS ray-path coverage obtained from the IGS network (right, blue) and the full IGS, AFREF and UNAVCO network (left, red) during a typical imaging period (0:00 – 7:30 UT on 3 December 2012). 300 km ray-path pierce points are shown in black. Observations represented here are used in the 4:00 UT, 3 December 2012 images.

Figure 14 shows that the full network provides far more ray-path coverage than the IGS network alone. The IGS network has only isolated patches of coverage during the test period, whilst the full network only has a few large gaps. There are numerous redundant receivers in the full network – a similar level of coverage could be achieved with far fewer receivers, but such a network would be vulnerable to station outages.

5.3.3 Case studies

It is useful to examine images from the two networks alongside the original IRI simulations in order to understand how the structures in the ionosphere affect image accuracy. Figure 16 shows a series of case studies that depict the way the image quality varies depending on the ionospheric state. Examples were selected to show the features that occur at different times of day.

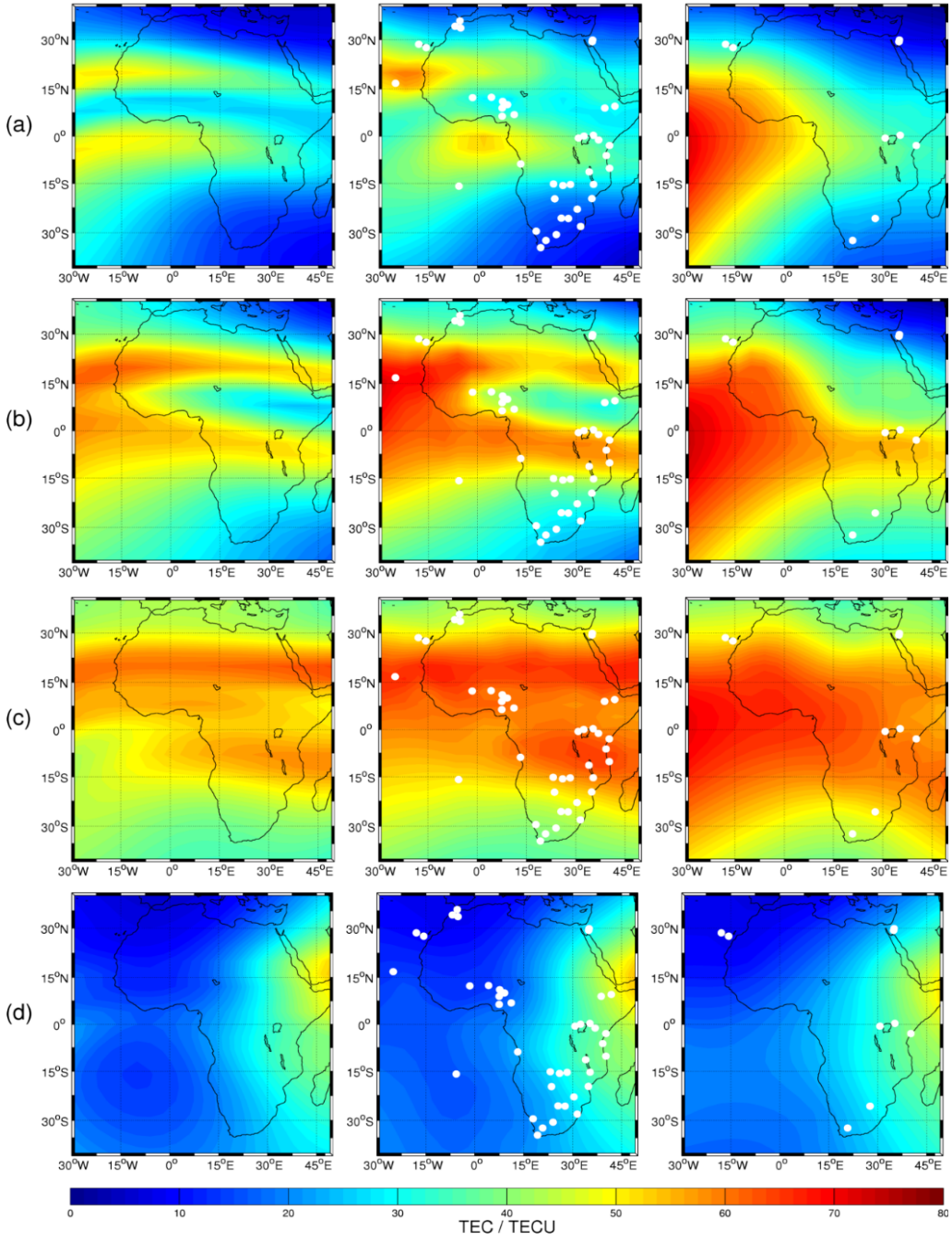

Figure 15: IRI simulations (left), reconstructions based on all available data (middle), reconstructions based on just IGS data (right). (a) - 22:00 UT on 2 December 2012. (b) - 17:00 UT on 3 December 2012. (c) - 12:00 UT on 7 December 2012. (d) - 05:00 UT on 9 December 2012. GPS sites in white.

With the exception of the night-time / early morning case study, (d), all the IGS-only images (shown in the right column of Figure 16) overestimate the IRI simulated 'truth' (shown in the left column) more than the images based on the extended network of data. However, the errors are generally overestimates in both cases. Both networks perform well in case study (d) because of the lack of large-scale ionospheric structures.

The different case studies give an insight into the reasons for the overestimation. Case study (a) shows a significant overestimation of TEC in the western region of the IGS reconstruction. This is caused by the regularization condition, which extrapolates the gradients observed in the north and east of the image. The lack of data in the western region allows the TEC values to continue increasing until the edge of the image. This does not occur in the reconstruction based on all the available data because there are numerous active receiver sites in the western part of the image. The lack of data could equally have resulted in artificially low TEC values, but the distribution of receiver sites in this case means that TEC generally increases from observed to unobserved parts of the grid. This positive gradient causes the artificial enhancements observed.

The IGS-only reconstruction in case study (b) has the western TEC enhancement most clearly shown in case study (a), but study (b) also contains a more unusual artefact. In this case, the IGS-only reconstruction has underestimated the northern band of ionization caused by the equatorial ionization anomaly in the north-eastern sector. It appears the sites above and below this phenomenon have measured only small positive gradients towards the band of ionization and therefore the interpolation has resulted in an underestimate. The reconstruction with all the available data shows that it is possible to image this phenomenon accurately when two east African receivers are present.

In case study (c) and to a lesser extent (b), both reconstructed images overestimate the TEC values of the IRI 'truth' image. The overestimation is caused by the regularization condition, which extrapolates a constant gradient across data-sparse regions. This is evident in the west of the images, where there are few observations available.

5.3.4 RMS Errors

To measure the accuracy of the images, we calculate the differences between the vertical TEC from the images and the vertical TEC from the simulations that the images are based on. Spatially distributed Root-Mean-Square (RMS) errors are calculated based on the errors of the entire period (2-16 December 2012) and plotted in Figure 16.

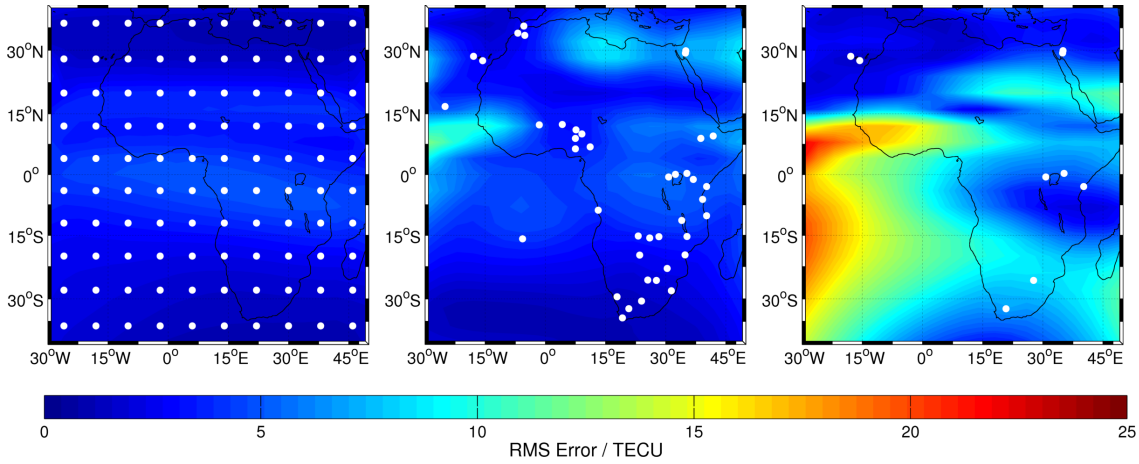

Figure 16: The RMS errors of the three sets of simulated reconstructions. On the left are the RMS errors of the reconstructions based on the simulated network, in the centre are the errors from the 'full' network and on the right are the errors of the reconstructions based on just the IGS receivers. The receiver sites used in each set of reconstructions are marked in white.

The results show that the reconstructions based on more data have far lower errors. The overall RMS error figures are 3.0 TECU for the fictional, ideal network, 4.5 TECU for the "full" network and 9.5 TECU for the IGS-only network. The errors are generally larger away from the receiver sites, as is shown in Figure 16. The IGS-only images had very large RMS errors (around 15 TECU RMS) in the region south of 15° N and west of 15° E. This is because there were no observations from IGS receivers in this region during the test period.

5.3.5 Images based on real observations

The results of Sections 5.3.3 and 5.3.4 show that it is possible to get reasonably accurate images of the ionosphere over the mainland area of Africa with the 'full' (IGS, AFREF and UNAVCO) receiver network. In this section, images based on real data are shown. The case studies selected here are at the same times are those of

Section 5.3.3. The images are shown in Figure 17.

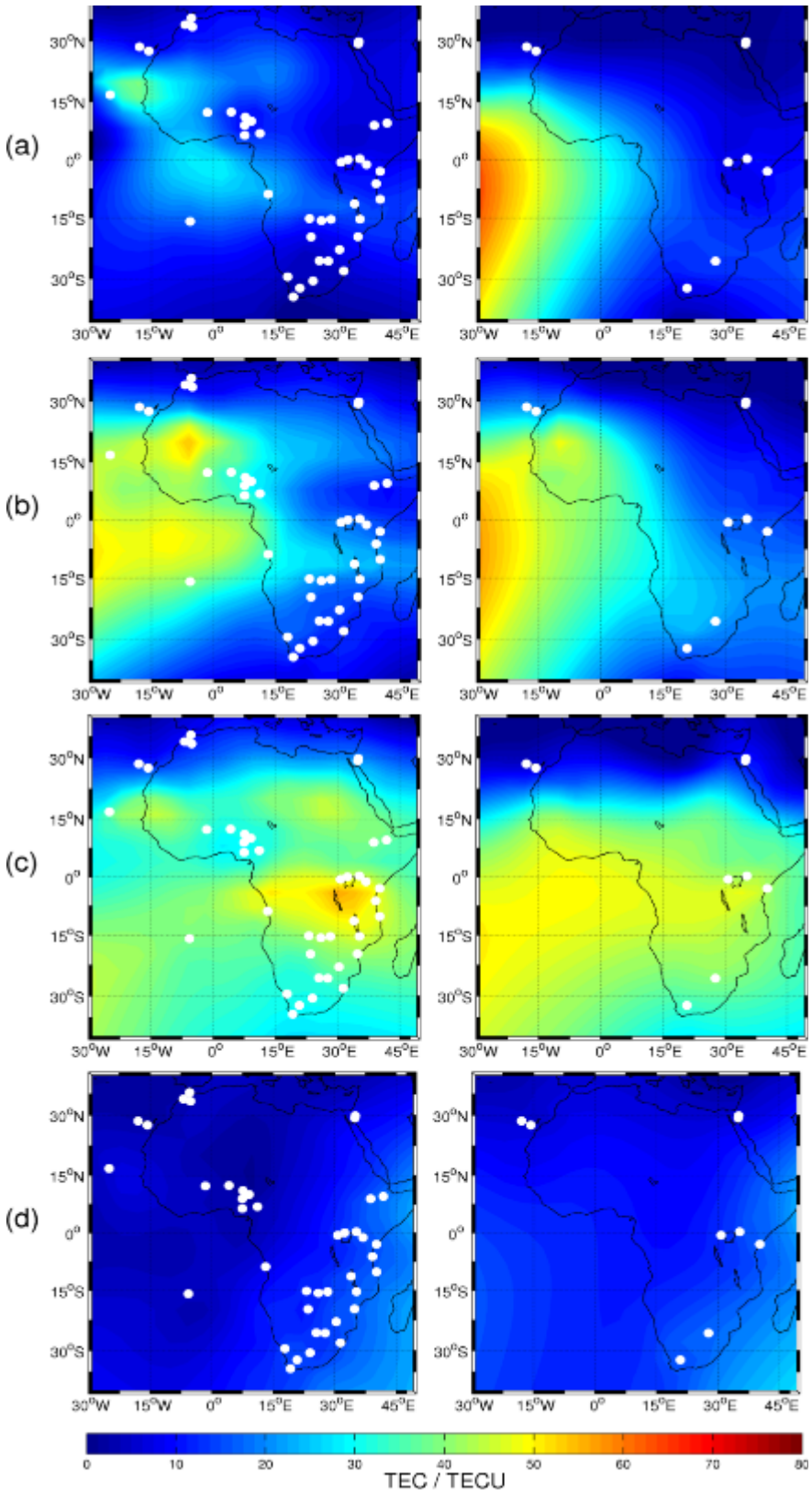

Figure 17: Reconstructions based on all available data (left) and reconstructions based on just IGS data (right). Row (a) is for 22:00 UT on 2 December 2012. Row (b) is for 17:00 UT on 3 December 2012. Row (c) is for 12:00 UT on 7 December 2012. Row (d) is for 05:00 UT on 9 December 2012.

The images in Figure 17 correspond to the case studies shown in Figure 16. In general, TEC values are lower in Figure 17. This is not surprising – the IRI simulated ionosphere is only intended to provide a monthly median specification. The images in Figure 17 also appear to have some of the same artefacts described in Section 5.3.4. The images based on only IGS data in rows (a) and (b) have large western TEC enhancements because of lack of data. The images based on all the data in (a) and (b) clearly show the two latitudinal bands of ionization associated with the equatorial ionization anomaly, a feature that is absent from the IGS-only images. Case study (c) in Figure 16 showed a clear overestimation of TEC values compared with IRI's 'truth'. In the absence of an independent 'truth', it is impossible to say whether the images in row (c) of Figure 17 overestimate or underestimate the true TEC values present in the ionosphere. Case study (c) does show that it is possible to produce reasonably good images from just the IGS receiver data - both images have similar features and absolute values across the land area. In this case, the real ionospheric conditions matched the inversion regularization condition closely enough that it was possible to extrapolate the available data accurately. As was seen in the simulation results of Figure 16, there is good agreement between the images from the two networks in case study (d). However, the westward enhancement (caused by lack of data) is visible over the Atlantic in the IGS-only image.

5.4 Discussion

Accurate ionospheric specification is necessary for improving high-frequency communications. The aim of this study was to determine how much GPS receiver coverage is required to accurately specify the ionosphere over Africa. Figure 16, presented in Section 5.3.4, shows the RMS errors of two sets of images, which are based on an IRI simulated 'truth' ionosphere. The results show that it is possible to achieve below 5 TECU RMS error over most of Africa if data from the full network is present. The 'full' network is the combined output of the IGS, UNAVCO and AFREF stations. This RMS error estimate takes into account imaging problems caused by sub-grid scale structures (see Section 5.2.3), but assumes that real large-scale ionospheric structures are similar to those in IRI-2012. Most of the full network stations are from a short-term campaign, so the IGS-only results represent the accuracy that can currently be achieved. In this case, RMS errors are often well

above 10 TECU over Africa. The situation is even worse over the ocean, where RMS errors from the IGS-only simulations exceed 20 TECU. These results make a clear case for the addition of more permanent operational receiver sites in Africa.

Whilst there is a clear need for more receiver coverage, care must be taken in siting the receivers to maximize their effectiveness. Figure 14 shows the ray-path coverage obtained from the two networks. In the full network, numerous receivers are grouped together. The observations from these clustered receivers provide little more ionospheric information than would be provided by a single receiver in the centre of the cluster. The same number of receivers could specify the ionosphere far more accurately if the receivers were evenly distributed across the grid. Alternatively, the same image quality could be achieved using a far smaller number of receivers. However, there is an important benefit to having clusters of receivers that is not obvious from these results. The GPS receiver sites frequently experience extended data gaps – in fact the IGS network has a number of receivers that produced no usable data during the test period. The creation of multiple receivers in close proximity to each other should be encouraged in areas where accurate ionospheric specification is required.

When considering the potential deployment of new GPS receivers for the purpose of ionospheric specification, it is useful to have an idea of the maximum achievable accuracy possible through ground-based GPS tomography. The results presented in Section 5.3.1 give an estimate of this upper limit by producing images from a dense, fictitious network of receivers. The RMS errors of these images (shown in Figure 13) are lower than those presented in Figure 16 in Section 5.3.4, which is as expected since those results are based on limited and unevenly distributed observations. However, it should not be assumed that perfect images could be produced if only we had enough GPS receivers. As is shown in Figure 13, RMS errors of up to 5 TECU are still present in images based on a dense network. These errors are partly due to the inaccuracies of the assumptions used in the imaging process. It is not always possible to reproduce the true ionosphere by a linear combination of two vertical basis functions, and the ionosphere does not always match the specified regularization condition. For example, *Zapfe et al.* [2007] achieved optimal TEC image accuracy by using four vertical basis functions. In general a greater number of basis functions provides the inversion with more degrees of freedom. This might

improve the image accuracy if there are enough observations to constrain the solution, but could otherwise lead to instability.

Different GNSS systems may also be utilized to improve ionospheric image accuracy. In situations where there are less usable satellites than usable receivers, as is the case with nearly all ionospheric imaging, image quality will normally benefit more from the addition of satellites than from the addition of receivers. This is because the new satellites will create more new ray-paths (one to each visible receiver) than a new receiver would (one to each visible satellite). Bearing this in mind, any new receivers installed should be designed to receive data from as many different GNSS networks as possible. Receivers placed in coverage gaps will clearly improve imaging accuracy far more than those placed close to existing receivers. The inclusion of observations that provide detailed information on the vertical electron density profile, such as observations from ionosondes or radio occultation measurements, would also provide significant improvements to ionospheric imaging accuracy, as is shown in *Chartier et al.* [2012b].

In any realistic situation, the number and type of observations is likely to be insufficient to specify the ionosphere without making any assumptions. For this reason, it is necessary to use techniques such as regularization and basis function transformations. In Section 5.3.3, several image artefacts were highlighted and attributed to the regularization condition used here (which favours zero second derivatives in the horizontal directions and in time). The most pronounced error was the TEC enhancement in the south-west of the images. This problem was caused by the lack of data in that region. It would be possible to overcome this problem by using different imaging techniques. For example, a background model term could be included in the inversion. A background model was not used here because such an approach could introduce significant biases across the image.

The results in Section 5.3.5 show that it is sometimes possible to produce reasonably good images from very few observations. In particular, case study (c) of Figure 17 showed very similar images produced from the IGS and 'full' networks. This was possible because the inversion assumptions (regularization and basis functions) matched the real ionospheric conditions quite well at that time. Case studies (a) and (b) of Figure 17 showed that this is often not the case – at least the inversion

assumptions are seldom sufficient to provide good image accuracy in the absence of sufficient data.

In summary, the results of this chapter show that African ionospheric images can be made significantly more accurate if additional receivers (beyond those available through the IGS network) are used. The use of these additional receivers can reduce RMS TEC errors by a factor of two over a large part of Africa. The results suggests that any new operational receivers should be deployed far from currently operating receivers. Efforts should also be made to increase the reliability of the existing network so that network redundancy is not required. Both new and existing receivers should be made capable of using as many different GNSS networks as possible. This study was focussed on reconstructing horizontal structures. In the next chapter, a new approach to imaging the vertical electron density distribution will be described.

6 Ionosondes in GNSS tomography

Abstract

In this chapter, the problem of resolving vertical electron density structures with predominantly high elevation angle rays from ground-based GPS receivers is addressed. HF radio communications depend on accurate specification of the ionospheric vertical profile. Observations from the Jicamarca incoherent scatter radar are used as a ground truth with which to validate a new method of incorporating ionosonde observations into GPS ionospheric tomography. The method uses ionosonde measurements to set vertical basis functions adaptively and then also ingests the observations directly. The result is that the accuracy of the vertical electron density profile is greatly improved. This technique uses auto-scaled ionosonde observations and is therefore suitable for use in near-real-time applications.

The results of this chapter were published in *Chartier et al.* [2012b].

6.1 Introduction

The aim of this study is to investigate the use of ionosonde observations within GPS tomographic imaging of the ionosphere. Improved imaging of the vertical structure of the ionosphere could lead to better scientific understanding of this region. Applications such as radio signal ray-tracing would benefit greatly from improved accuracy in the vertical structure of electron density reconstructions. In addition, the accurate specification of electron density is of interest for the constraint of the ionospheric portion of coupled physical models.

6.1.1 Physical processes and the vertical structure

Various aspects of research into ionospheric physical processes depend on

observations of the vertical profile. For example, the *Yin et al.* [2006] study of the F-region uplift during storm-time showed a plasma uplift with latitudinal and longitudinal dependence. The authors highlighted a limitation of ionosonde observations for this application, namely that they give no information on the topside plasma distribution. *Tsurutani et al.* [2004] described a global plasma uplift caused by a shock from the interplanetary magnetic field, which resulted in a TEC increase of 80% at mid-latitudes. Accurate representation of the ionospheric vertical structure will aid further study of this important phenomenon. The work presented here is also important for the development of ionospheric and coupled modeling efforts. A modelling study by *Lin et al.* [2009] showed the creation of an additional ionospheric layer in the equatorial region during a storm-time uplift. The authors linked the phenomenon with meridional neutral winds crossing the equatorial boundary into the opposite hemisphere. The work of *Jin et al.* [2011] on the Ground-to-topside model of Atmosphere and Ionosphere for Aeronomy (GAIA) successfully reproduced a four-peak structure in the daytime equatorial ionization anomaly, showing that non-migrating tides in the troposphere are responsible for the phenomenon. The quality of vertical electron density profile observations is a topic of current research. For example, a recent study by *Ely et al.* [2012] sought to test the quality of current vertical profile observations by comparing GPS radio occultation profiles with profiles from ionosondes in Brazil. There have also been several studies [e.g. *Abdullah et al.*, 2010, *Azzarone et al.*, 2012; *Tsai et al.*, 2010; *Warrington et al.*, 2012] on ray-tracing through the ionosphere, an application that depends strongly on the vertical electron density structure. The technique presented in this chapter could be used to improve the vertical structure of near-real-time electron density images, which would be beneficial for operational ray-tracing applications.

6.1.2 The new technique

Whilst GPS-derived tomography can produce good total electron content maps, the rays do not contain sufficient information to accurately specify the vertical profile. Ionosondes provide observations of the vertical profile that could be used to improve the GPS images. The approach evaluated here assimilates ionosonde observations directly by treating the ionosonde peak density estimate as a measurement point and defining the vertical profile observations as a series of gradients away from that peak. This approach is more flexible than defining each point in absolute terms, as it

allows the absolute values to change whilst retaining the shape. In addition, vertical basis functions, used to constrain the vertical profile in the inversion, are set adaptively using the ionosonde data. A study was performed in the equatorial region of South America. Results are validated with data from the Jicamarca incoherent scatter radar and with independent GPS data. Previous work has investigated the use of ionosonde data in electron density tomography at mid-latitudes [*Garcia-Fernandez et al.*, 2003; *Dear and Mitchell*, 2007] and high latitudes [*Kersley et al.*, 1993] but this is the first study at low-latitudes. This is important because the vertical electron density profile is harder to image at low-latitudes due to the GPS satellite geometry. A technique by *Stankov et al.* [2011] produces local near-real-time electron density profiles by combining ground-based GNSS and ionosonde measurements. However, the present study is the first to show a technique that simultaneously assimilates ionosonde observations and also uses adaptive basis functions set by those ionosonde observations. The technique presented here improves the resolution of the images by including additional observations.

Since the technique presented in this chapter uses observations from an ionosonde and the results are validated with Jicamarca incoherent scatter radar observations, background information on those two instruments will be given in Sections 6.1.3 and 6.1.4.

6.1.3 The digital ionosonde

Ionosondes sound out a profile of electron density up to the F region peak. They provide electron density profiles by sending electromagnetic waves of frequencies up to tens of MHz towards the ionosphere. Echoes of the wave are detected by a receiver in the vicinity of the transmitter. The virtual height is estimated from the time it takes to receive the echo, whilst the electron density is derived from the reflected frequency. Assuming the electron thermal speed is negligible, the plasma frequency is proportional to the square root of the electron density. *Cannon et al.* [1992] describe a technique for measuring ionospheric drift using Doppler interferometry. *Bibl and Reinisch* [1978] state that the digital ionosonde, or 'digisonde', can be used for both ionospheric monitoring and research. The low cost of the instrument allows a dense network of sounders to exist. Digisondes provide information on ionospheric drift as well as electron density. *Reinisch et al.* [2004]

report that a global network of over 70 digisondes exists and that auto-scaling software, such as the Automatic Real-Time Ionogram Scaler with True height (ARTIST) by *Reinisch and Huang* [1983] can provide estimates of ionospheric parameters at these key locations. Autoscala, by *Pezzopane and Scotto* [2004], is another software tool for automatically scaling ionograms. *Pezzopane and Scotto* [2007] found Autoscala and ARTIST 4.5 generated parameters both generally agreed with manual scaling. ARTIST 4.5 generated invalid data for manually unscalable ionograms whereas Autoscala usually gave no result in those circumstances. *Galkin and Reinisch* [2008] report that poorly scaled ionograms are given a low confidence estimate in the newer ARTIST 5 software and that the auto-scaled *foF2* agrees with the manually-scaled value to within 0.3 MHz for 95% of the test period. Digisonde data has been used for operational modelling efforts including the Parameterised Real-Time Ionospheric Specification Model (PRISM) and USU GAIM projects [*Daniell et al.*, 1995; *Sojka et al.*, 2003]. The real-time availability of these observations means they could be useful for other assimilation schemes in the future.

6.1.4 Jicamarca incoherent scatter radar

The Jicamarca instrument is the world's largest incoherent scatter radar. The radar can provide absolute electron density observations without scaling the peak to an ionosonde measurement of *NmF2*. This is because the instrument is capable of measuring *Faraday* rotation and the magnetic dip angle at Jicamarca is about one degree, so the beam can be aimed perpendicular to the magnetic field. *Farley* [1969] describes the method used to measure electron density at Jicamarca. The radar's beam is set a few degrees off perpendicular to the magnetic field because the power spectrum of incoherently scattered magnetic waves becomes sharper when the beam is close to orthogonal (*Farley et al.* [1961]). The phase shift is equal to twice the rotation of the plane of polarization of a linearly polarized wave, since the wave must pass twice through the ionosphere. *Hysell et al.* [2008] give further details of the unique features of the Jicamarca incoherent scatter radar.

6.1.5 Ionosondes in GPS tomography

Kersley et al. [1993] explored the possibility of using ionosondes in a tomographic

imaging technique. They created several background ionospheres from the IRI-90 empirical model to match observed ionosonde peak height and density at different times, then used those model backgrounds in their inversion. They found the ionosonde measurements they used sometimes overestimated peak density by 70 % compared to the EISCAT incoherent scatter radar, but suggested the errors were due to particularly steep electron density gradients in the region. The authors found the lower portion of their images were improved by a factor of two using the ionosonde input, but that the topside was adversely affected by the known overestimation of topside densities in IRI-90. *Ma et al.* [2005] used GPS and ionosondes to train a neural network. The network training was carried out by minimizing the squared residuals of an integral equation. The authors found that ionosonde data was very useful in improving the vertical profile of their images. *Yin and Mitchell* [2005] used ionosonde observations to create adaptive basis functions in an earlier version of MIDAS. IDA, by *Bust et al.* [2000], is capable of directly ingesting ionosonde electron density observations into a three-dimensional model grid, whilst *Galkin et al.* [2012] described a new method for real-time assimilation of ionosonde measurements into IRI. *Pezzopane et al.* [2011] showed an optimal interpolation method for assimilating ionosonde observations into regional and global empirical models, whilst *McNamara et al.* [2011] described a Gauss-Markov Kalman filter approach to assimilating ionosonde observations in the USU GAIM scheme (validated in *McNamara et al.* [2008]), though they did not find any improvement in the accuracy of their *foF2* estimates. The technique presented in this chapter differs from previous work in that it directly assimilates ionosonde observations into a basis function-decomposed space, where the basis functions are set adaptively according to the same ionosonde observations.

6.1.6 Ionospheric imaging

The ionospheric imaging algorithm used in this study, known as MIDAS, is a three-dimensional, time-dependent algorithm for imaging the ionosphere using multiple data sources including GPS phase data. MIDAS is described in detail in Section 3.7. Similar techniques have been developed by *Bust et al.* [2000, 2004], whilst other data assimilation methods are described by *Schunk et al.* [2004], *Mandrake et al.* [2005] and *Angling and Jackson-Booth* [2011]. Due to the geometry of the ray paths and the generally poor quality of low elevation rays, GPS-derived tomographic images often

contain little information on the vertical distribution of plasma. This is a particular problem in low latitude regions, such as around Jicamarca, due to the GPS ray-path geometry.

6.2 Method

A new technique was developed for including ionosonde observations in GPS tomography. This was compared with incoherent scatter profiles from the same location as the ionosonde and was also tested against calibrated GPS slant TEC observations from the surrounding area. A six day period from 8 to 14 July 2008 was selected for the experiment as incoherent scatter data were available at those times. *Kp* reached a maximum of 4+ during this period, whilst *F10.7* was around 67-68 throughout. A MIDAS run with no ionosonde data was performed as a control. A rotated grid was used to accommodate the shape of northern South America without including ocean-only cells with no GPS receiver sites. This grid and the receiver sites used for this experiment are shown in Figure 18. The figure also shows resolution masked TEC maps. This masking technique is described in the Appendix.

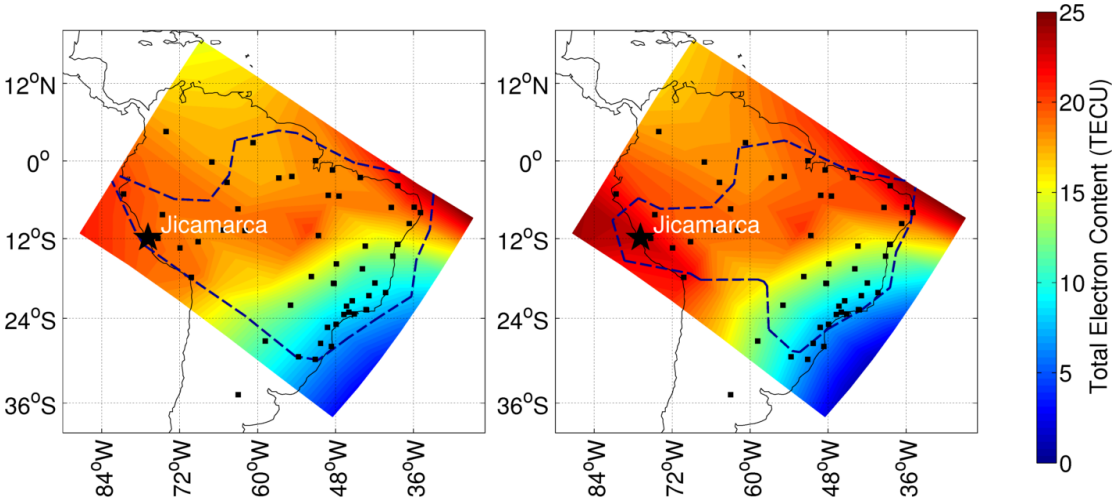

Figure 18: A TEC reconstruction using the new technique on the right (with ionosonde observations above Jicamarca and ionosonde-derived basis functions throughout) and the standard technique on the left (using only GPS observations and basis functions derived from IRI). The dashed contour line encloses the section of the image identified as most reliable by the resolution mapping algorithm. GPS receiver sites used in the inversion are shown in blue. The reconstructions show 14:00 UT (or 10:00 in Lima, Peru) on 11 July 2008.

6.2.1 Ionosonde observations in GPS tomography

Ionosonde observations can only be available in near-real-time - a requirement for inclusion in a data assimilation scheme - if auto-scaling software is used. ARTIST 4.5, described by *Reinisch et al.* [2005], was used here to provide the data shown in Figure 15 (b). The software package creates an ionospheric profile from auto-scaled parameters such as virtual height and critical frequency, which are calculated from the raw ionogram. On comparison with the incoherent scatter profiles (shown in Figure 19 (a)), it is clear that the ARTIST topside has significant errors. In order to estimate the topside scale height more accurately, IRI 2007 [*Bilitza and Reinisch,* 2008] is run with ionosonde peak height and density inputs. The scale height is calculated from the vertical electron density profile produced by IRI, which is run with the standard topside rather than the option based on the NeQuick model by *Nava et al.* [2009]. The more recent Vary-Chap approach, developed by *Nsumei et al.* [2012], could be used in a future version of the technique.

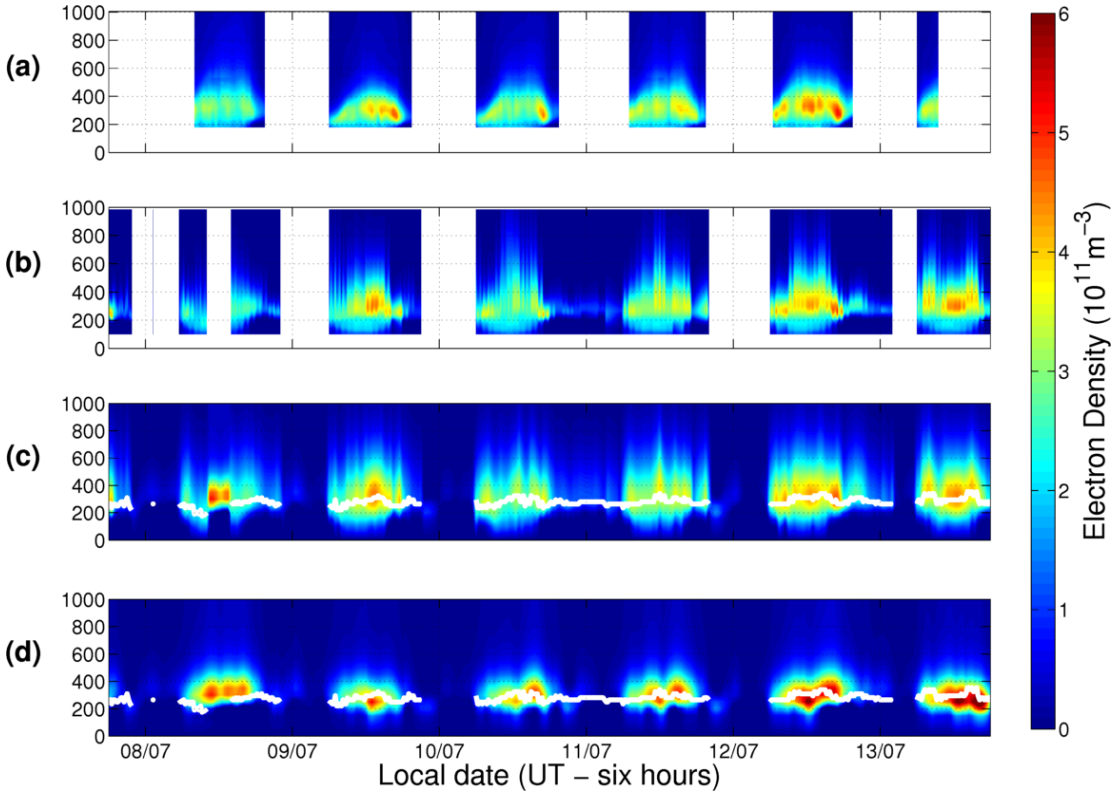

Figure 19: (a) Jicamarca incoherent scatter observations (ground truth), (b) Jicamarca ionospheric profiles derived from ionosonde observations, (c) a MIDAS inversion that included the ionosonde observations, and (d) a MIDAS inversion that did not have the ionosonde observations. White lines show the ionosonde peak height.

Appropriate basis functions must be selected in order for ionosonde peak height and density observations to be successfully incorporated into GPS-derived tomographic images. A poor choice of basis functions would mean that the measured ionosonde peak height could not be reproduced, which would lead to a large overestimation of *NmF2*. This is because an electron density value from above or below the peak of the basis function profile would be scaled to the observed *NmF2* and that would scale up the whole profile. The alternative, that the profile would be scaled down, is not possible because the values from above or below the basis function's *hmF2* must always be lower than the basis function's peak density. A technique was implemented to create basis functions adaptively so that ionosonde observations could be successfully assimilated at each time-step. The adaptive basis function technique shown here is theoretically similar to the work of *Materassi and Mitchell* [2005a], where ionosonde observations are used to set parameters in the creation of basis functions. In the technique described here, this method is used so that the inversion can match the peak and scale heights specified by the directly assimilated ionosonde observations, rather than simply to provide more realistic basis functions. As mentioned above, the basis functions were based on profiles, derived from Chapman's equations, that matched the ionosonde peak and bottom-side scale heights. A topside scale height was derived by running IRI 2007 with the observed *hmF2* and *NmF2* from the ionosonde. This was found to produce better agreement with the incoherent scatter radar observations than relying directly on the ARTIST topside for an estimate of topside scale height. In our work the following relation, derived from the work of *Chapman* [1931], was used to calculate an electron density profile:

$$N_e \propto \exp(1 - \zeta - \exp(- \zeta)) \tag{37}$$

where N_e is the electron density at a point and ζ is defined as:

$$\zeta = (h - hmF2) / H \tag{38}$$

Here, h is the height of the point above the ground, *hmF2* is the peak height and H is the scale height. Of course, Equation 37 does not give realistic electron density values, but we are simply interested in producing reasonable profile shapes from which to create normalised basis functions.

The most straightforward method to assimilate an electron density profile from an ionosonde would be to express the profile as a series of point electron densities in the observation vector, $\boldsymbol{b}$ (see Equation 26 in Section 3.4). There are two problems with this method. Firstly, GPS ionospheric tomography systems use slant TEC observations rather than electron densities. This can be overcome by converting the electron densities to a series of short horizontal TEC line segments. The second and more serious issue is that only the peak density is accurately observed by the ionosonde. Both the measured bottom-side and the extrapolated topside are likely to contain significant biases, a problem which is exacerbated by auto-scaling errors. Direct assimilation of these biased observations can create artefacts because the total slant TEC observed by the GPS rays must be accommodated somewhere along the ray paths. One way to overcome this limitation would be to assimilate only the peak value, but it was found that this method did not sufficiently constrain the shape of the profile. Instead, we developed an approach that kept the shape information present in the profile, but allowed the absolute values to vary in order to mitigate systematic biases away from the peak. The ionosonde peak density was treated as an absolute point measurement, but the rest of the profile was treated as a series of gradients away from that peak value. These gradients were calculated by taking the differences between adjacent profile points. It was found that the inclusion of profile gradient information lead to a slight improvement in the vertical profile compared with inversions using only peak height and density. This was because information on the shape of the profile was preserved without forcing the absolute values.

The final issue to be considered for the assimilation of ionosonde observations into a GPS-derived tomographic system was whether to use auto-scaled or manually-scaled observations. It is desirable to use auto-scaled ionosonde observations so that the technique can be used operationally. Auto-scaling also reduces the amount of labour required for historical studies. However, it is widely accepted that manually scaled observations are more accurate. The quality of auto-scaled ionosonde observations was assessed to see whether it would be feasible to use them. Ionosonde *NmF2* and *hmF2* were compared with equivalent measurements from the incoherent scatter radar. Mean, root-mean-square (RMS) and maximum errors were calculated (see Table 1) and showed that the ionosonde observations contained some significant discrepancies. However, the auto-scaled observations were still good enough to improve the accuracy of the technique compared with the control, as was shown in

the validation (discussed in more detail in Section 6.3). During the development of the algorithm, a preliminary experiment was run with the incoherent scatter radar observations used for direct assimilation. That experiment predictably produced very close agreement with the raw incoherent scatter observations, but also resulted in slightly improved performance compared with GPS slant TEC observations in the region. The preliminary experiment showed how the technique could perform with very well auto-scaled ionosonde observations.

	NmF2 (10^{11} m^{-3})	*hmF2* (km)
Mean error	-0.12	19.7
RMS error	0.51	39.8
Maximum absolute error	2.11	155

Table 1: The mean, RMS and maximum absolute error of ARTIST 4.5 auto-scaled ionograms, compared with Jicamarca incoherent scatter radar observations at the same location. Typical daytime values of NmF2 are around 5 x 10^{11} m^{-3}, whilst hmF2 values are usually between 200 and 400 km.

6.2.2 Ionospheric images

GPS phase observations were gathered over a five-hour time window. Rays from the same satellite-receiver pair were differenced to give TEC gradients in space and time. A 'best fit' solution for electron densities was found at 30-minute intervals through the time window using a least-squares technique, and the solution for the centre of the time window was selected. A regularisation condition suppressing changes in gradient in time and space was included. The two MIDAS inversions, one that included ionosonde information and one control run, use the same GPS data on the same rotated 5 degrees latitude by 10 degrees longitude grid, shown in Figure 18. A grid with smaller voxels would have been preferable for imaging small-scale structures, but was found to offer no advantage due to the lack of GPS data coverage in the region. GPS observations contain information on the plasmasphere that must be accommodated somewhere in the inversion, so the grid top was set to 2000 km. Figure 19 shows only 0 - 1000 km for easier comparison with the incoherent scatter observations. Both runs used two basis functions to constrain the vertical profile at each latitude/longitude grid point. In the control run, these were derived from IRI. The new technique used the same basis functions when ionosonde data was

unavailable, but set the basis functions at each 30-minute timestep when ionosonde observations were available. Reconstructions were produced between 8 and 14 July 2008.

6.2.3 Validation with GPS slant TEC observations

Independent GPS slant TEC from phase-leveled, bias corrected pseudorange observations were used to check the quality of the two MIDAS runs. *Choi et al.* [2011] and *Wilson and Mannucci* [1993] state that pseudo-range contains an absolute but noisy measurement of slant TEC including significant hardware biases. In this case, the biases were estimated at Jicamarca and noise was smoothed using the phase observations. Slant TEC from this GPS phase-levelling process was subtracted from the equivalent paths traced through the MIDAS image in order to estimate the error in the image. Five receiver sites were held back from the inversion so that they could be used in this test. Rays below 15 degrees elevation were excluded. A resolution mapping algorithm was used to mask out grid cells with little ray coverage so that spurious parts of the image did not disturb the results. The mask was based on a standard basis function for both tests, so any differences in coverage are due to the additional observations present in the new technique. The results of this test are shown in Table 2. The resolution mapping technique is described fully in the appendix. The effects of the resolution mask can be seen in Figure 18.

Run Type	Mean error (TECU)	RMS error (TECU)	No. of Rays
Adaptive basis functions + ionosonde peak and profile (from IRI)	-0.36	3.55	475
Fixed IRI-derived basis functions	0.64	4.02	468

Table 2: Mean and RMS errors of MIDAS images are compared with independent phase-levelled, bias-corrected GPS slant TEC observations from rays with elevations greater than 15 degrees. Both sets of images are masked according to data resolution, so only well-defined regions of the grid are tested. The new technique results in a slightly better conditioned problem due to the extra data used, which is why it has more valid rays.

6.3 Results

6.3.1 Vertical profile

Figure 19 shows that the ionosonde data significantly improved the vertical profile of the MIDAS images. An example of this improvement is shown in Figure 20.

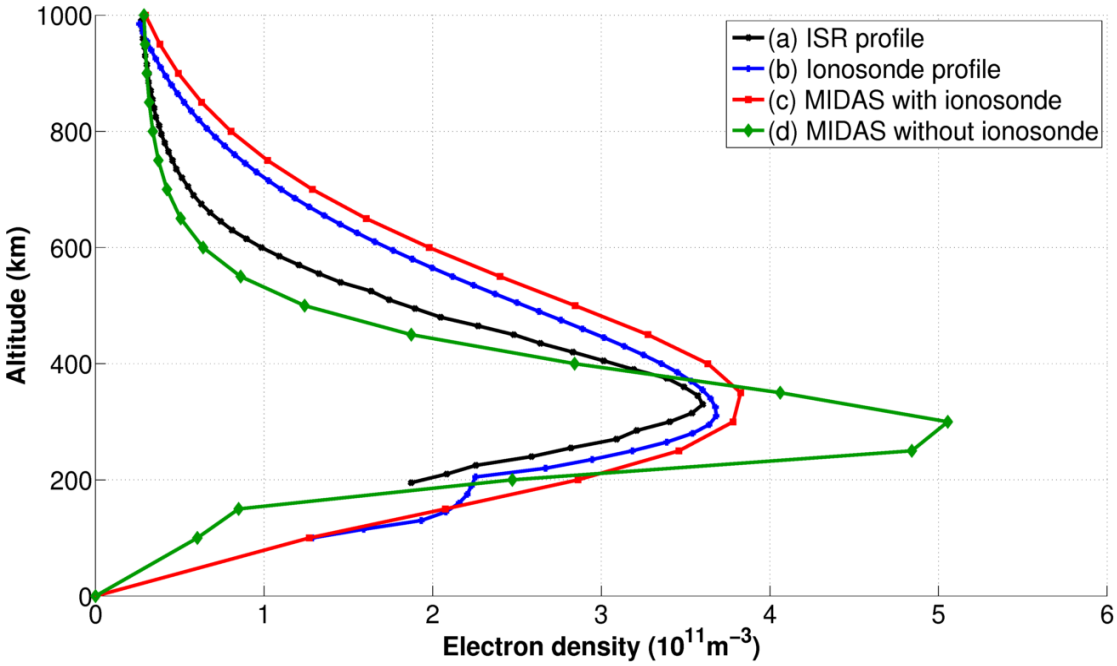

Figure 20: Vertical electron density profiles from the Jicamarca incoherent scatter radar, an auto-scaled profile from the Jicamarca ionosonde, a MIDAS inversion that included the auto-scaled profile, and a MIDAS inversion that did not include the auto-scaled profile. The Jicamarca incoherent scatter radar should be considered the ground truth here. The profiles are from 12 July 2008 at 17:00 UT (13:00 local time).

In general as well as in the example given in Figure 20 above, the run without ionosonde data (d) gave reasonable peak heights but significantly overestimated peak densities compared to the incoherent scatter radar. The run with ionosonde data (c) gave far more accurate profiles than (d) in comparison with the incoherent scatter radar, but predictably agreed even more closely with the ionosonde. In a numerical comparison with the incoherent scatter radar, the new technique gave mean *NmF2* error of -0.07 x 10^{11} electrons/m^3 compared with 0.27 x 10^{11} electrons/m^3 in the control, whilst Root-Mean-Square (RMS) error was 0.5 x 10^{11} electrons/m^3 compared with 1.0 x 10^{11} electrons/m^3 in the control. The new technique also resulted in 0.1 km mean error in *hmF2* compared with -3.9 km in the control run,

whilst RMS *hmF2* error was around 40 km in both cases. These large RMS errors in *hmF2* were due to the size of the voxels used in the reconstructions, which were 50 km in height extent and 5 degrees by 10 degrees in the rotated latitude and longitude coordinates. These had to be kept large due to the lack of data in some parts of the grid. The bottom portion of the incoherent scatter radar image was unusable, so it was not possible to assess the quality of the bottom-side scale height, but the new technique significantly overestimated the topside scale height. Figure 19 (b) shows that topside overestimation is present in the retrieved ionosonde profiles, so improvements in the auto-scaling and ionosonde inversion processing should further improve the performance of the technique. Even so, the lack of real topside observations must also be addressed. Some discontinuities were observed when the ionosonde was switched on or off, as the inversion reverted to IRI-derived basis functions. Many ionosondes produce continuous data, so this should not be a concern for the operational use of the technique.

6.3.2 TEC

Table 2 shows the results of comparing the MIDAS runs with independent, calibrated GPS slant TEC data from phase-levelled pseudo-range observations. These results show that the use of ionosonde observations decreased mean and RMS error compared with a MIDAS run that uses IRI-derived basis functions, giving -0.36 TECU mean error compared with 0.64 TECU in the control run and 3.55 TECU RMS compared with 4.02 TECU in the control run. This is because adaptively set basis functions can more accurately represent ionospheric conditions across a regional grid than a fixed set of basis functions for all times. The direct assimilation of ionosonde observations should also result in slightly improved slant TEC in the region. The number of GPS slant TEC observations available was slightly higher for the ionosonde run than for the control run. This was because of the resolution mapping algorithm, which excluded cells with poor ray-path coverage. The extra information from the ionosonde improved the conditioning of the problem so that more cells were included in this test.

6.3.3 Anomalous ionosonde observations

While ionosonde observations returned by ARTIST 4.5 are generally of a good

standard, they do occasionally contain significantly incorrect *hmF2* and *NmF2* values. For example, the profile from Jicamarca at 15:00 UT on 8 July 2008 significantly underestimated both *hmF2* and *NmF2*. Figure 21 shows that ionosonde profile as well as the equivalent incoherent scatter radar profile. The inclusion of these poor quality observations in the inversion affects the quality of the final image, as is also shown in Figure 21. Without the ionosonde observations, the inversion relies on *a priori* basis functions that can cause similar problems. In the example shown, the inversion without ionosonde observations happened to have an accurate peak height, but the peak density was significantly overestimated. This shows that *NmF2* could not be accurately reproduced using GPS observations alone. Although the inversion with ionosonde observations shared the peak height underestimation of the ionosonde profile, the problem of peak density underestimation was somewhat mitigated. In this case, with particularly poor auto-scaled data, the new technique produced a profile with an artefact, but the profile created without the ionosonde observations also contained significant inaccuracies.

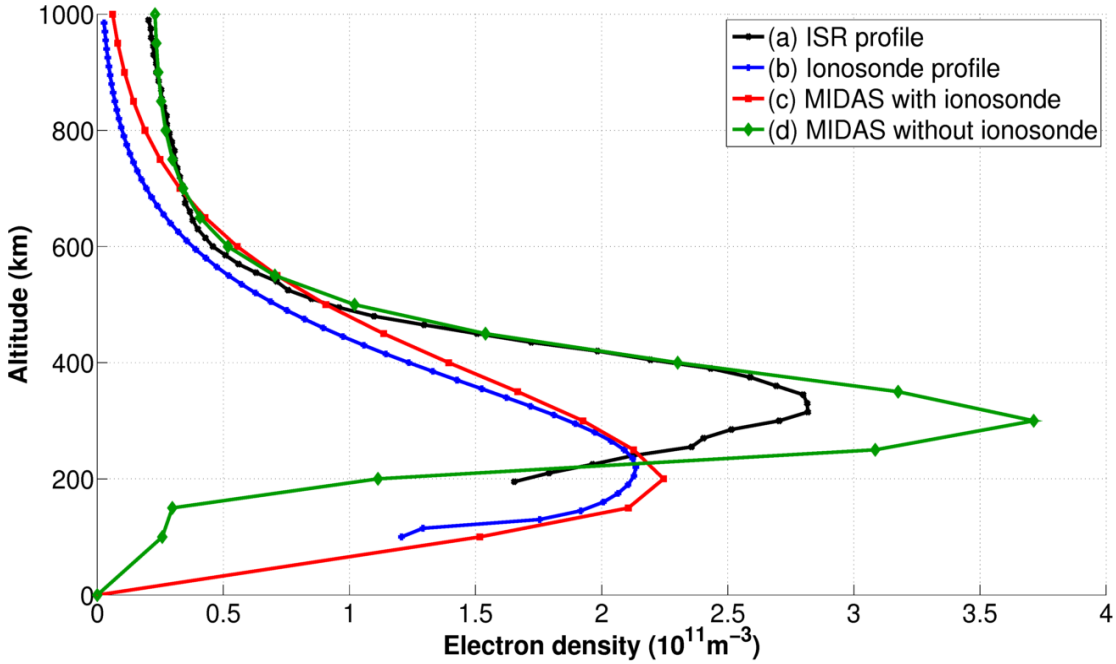

Figure 21: Vertical electron density profiles from the Jicamarca incoherent scatter radar, an auto-scaled profile from the Jicamarca ionosonde, a MIDAS inversion that included the auto-scaled profile, and a MIDAS inversion that did not include the auto-scaled profile. The Jicamarca incoherent scatter radar should be considered the ground truth here. The profiles are from 8 July 2008 at 15:00 UT (11:00 local time).

6.4 Discussion

This study shows that the vertical profile of GPS tomographic images can be improved when compared with independent incoherent scatter radar observations by the inclusion of auto-scaled ionosonde data. In principle, it follows that the inclusion of both additional observations and more realistic basis functions should result in improved vertical profiles and better slant TEC accuracy in the region. However, there is a risk that the inclusion of ionosonde observations with significant biases could cause artefacts and degrade the quality of the image. The problem of biases away from the peak was mitigated by including the vertical profile as a series of gradients. Our adaptive basis function scheme allowed observations of the peak to be easily accommodated by the inversion, and the technique included a regularisation term that smoothed biases caused by the ionosonde observations with GPS-derived TEC values from the region. However, the technique is still sensitive to very poor auto-scaled ionosonde observations. This highlights the continuing need for improvement of the ionosonde auto-scaling process.

A second benefit of the new technique is that it improves the overall quality of the image when compared with independent GPS slant TEC observations. The global ionosonde network will be useful for improving the quality of tomographic vertical profiles in the future. The addition of GPS observations in the inversion adds accuracy to the bottom-side ionosonde profiles, since it requires the solution to agree with observations in the region of the ionosonde. It should now be possible to distinguish between local structures, present only in the immediate vicinity of the ionosonde, and larger phenomena that will also be visible to GPS rays in the same grid cell as the ionosonde. Significant discrepancies between the ionosonde profile and the equivalent inverted profile may highlight the presence of a local structure, whilst close agreement may suggest a more widespread structure.

The problem of topside scale height overestimation is most likely due to topside overestimation in the input data, which may be corrected by improved topside modelling or, ideally, detailed observations of the topside. Currently, the basis functions reflect this overestimation of the topside scale height. Whilst it would be possible to create basis functions with smaller scale heights, it was found during

testing that using very sharp vertical profiles adversely affects TEC imaging in the region. This is because the inversion becomes far more sensitive to noise or representativeness errors in high-altitude observations. As it is, the technique will smooth out large-scale (present throughout the grid cell) topside overestimation from the ionosonde, since that will achieve the best fit with the GPS observations and the lower portion of the profile should be reasonably accurate. In regions of denser GPS ray-path coverage, it should be possible to reduce the size of the voxels, or volume elements, used in the inversion without creating instabilities. This would effectively reduce the representativeness error of the GPS rays and therefore allow more realistic basis functions to be used.

ARTIST 5, a newer ionogram auto-scaling and profile inversion software package, [*Galkin and Reinisch*, 2008] was not available for the Jicamarca ionosonde, but further work should test the performance of the new GPS tomographic inversion technique presented here with more advanced ionosonde auto-scaling techniques. In particular, the availability of error estimates in the newer software packages should allow users to overcome a significant weakness of the current technique, that large biases in the ionosondes can adversely affect the quality of the overall image. In terms of operational use, it would be preferable not to depend upon IRI 2007 to define the ionosonde topside scale height, since that model requires up-to-date estimates of the planetary magnetic disturbance index, *Kp,* and *F10.7*. It is possible that the Topside Sounder Model Profiler, discussed by *Kutiev et al.* [2009], could be used to reduce biases in the topside electron density profiles extrapolated from Digisonde observations.

The results support the conclusion that the new technique produces better vertical profiles and TEC reproductions than tomographic reconstructions based only on GPS observations with fixed basis functions, but the technique should be tested with more data and in a variety of geographic locations and geomagnetic conditions in future work.

A significant challenge for global implementation of the technique presented here is the question of using multiple ionosondes. This would mean changing the basis functions in a smooth manner across the grid. It would be desirable to use ionosonde-derived basis functions in the vicinity of the ionosonde and revert to basis

functions defined by IRI or a similar empirical model elsewhere. In that case, one would need to define appropriate ionospheric correlation lengths from which the region of influence of the ionosonde would be inferred. This definition of the correlation lengths is a long-standing issue in ionospheric imaging and data assimilation that requires further research.

So far, the results presented in this thesis have been aimed at real-time or historical specification of the ionosphere. The next chapter will deal with forecasting, which has the potential to complement specification efforts as well as being useful in its own right.

7 The effects of initializing different model fields on ionospheric forecasts

Abstract

In this chapter, the problem of forecasting the ionosphere is addressed. Forecasting could improve ionospheric specification in situations of intermittent observation coverage as well as being of use in mission planning. Ionospheric forecasting is challenging because the ionosphere is very strongly forced by other systems such as the thermosphere, the lower atmosphere, the Sun and the magnetosphere. The results shown here demonstrate that storm-time electron density forecast accuracy improvements of at least 10% can be achieved for over 19 hours by accurate initial specification of the thermospheric neutral composition, the most important component of which is the ratio of O to N_2. By contrast, initial specification of the ionosphere improves forecasts by 10% for less than four hours. The results also highlight the need for accurate specification of solar and geomagnetic drivers and modelling improvements by comparing model runs with real observations.

The results of this chapter were published in *Chartier et al.* [2013b].

7.1 Introduction

As is discussed in Chapter 1, many aspects of the thermosphere-ionosphere system affect human activities. Applications such as HF communications and satellite drag are considered in mission planning and so there is a need for forecasts of the thermosphere-ionosphere system. Forthcoming radio astronomy projects such as the Square Kilometre Array and the Low-Frequency Array will use transmission frequencies at which the ionosphere will be an important source of error. A forecast system that includes a dynamic model of the system could also help spread information from well-observed parts of the globe, thereby improving specification

in poorly-observed regions. However, thermosphere-ionosphere modelling is challenging because the system has a wide range of external drivers, such as insolation, particle precipitation, high-latitude electric fields and lower boundary forcings. In practice, many empirical and physics-based models represent climatological behaviour accurately but struggle to capture day-to-day variability. This is especially true in the case of storms. A data assimilation scheme could combine ionospheric and thermospheric observations with a model in order to create improved forecasts.

The term *data assimilation* refers to a series of techniques that adjust the current state estimates of models with the intention of improving forecasts. Assimilation techniques have been used successfully for this purpose in meteorology and other disciplines, but data assimilation is naturally most effective when the forecasts depend entirely on the initial values of the system. The ionosphere is a strongly forced system, especially during storms, so it is possible that any changes made to initial values by an assimilation scheme could quickly be overwhelmed by the effects of external driver changes. External drivers include lower boundary effects such as propagating gravity waves and tides, but during storms the major inputs are from solar and geomagnetic sources. For example the Interplanetary Magnetic Field (IMF) components B_x, B_y and B_z frequently exhibit significant changes during storms. The geomagnetic activity index, *Kp*, quantifies disturbances in the Earth's magnetic field that are caused by IMF disturbances. Large negative values of B_z drive strong convection and auroral precipitation in the high altitude ionosphere. These two processes cause enhanced Joule heating, which results in changes to the neutral composition and thermospheric neutral winds. Typically, these thermospheric changes lead to a decrease in high- and mid-latitude plasma densities. *Richmond and Lu*, [2000], describe the role of penetration electric fields in causing large increases in mid-latitude plasma densities during storms. The direct penetration of electric fields from the polar regions down to the equatorial latitudes could enhance the daytime eastward electric field. This increase in eastward electric field strength would enhance the equatorial or Appleton anomaly, causing the density maxima to move further poleward and the trough at the magnetic equator to be enhanced. The competing negative and positive effects on storm-time ionospheric densities cause each storm to be unique and difficult to forecast.

The aim of this study is to establish whether a data assimilation scheme can produce accurate ionospheric forecasts in the absence of good external driver estimates. A coupled thermosphere-ionosphere model, described in Section 7.1.1, is used for this study. The experiments test the storm-time forecasting performance of an assimilation scheme in which a run of the model is taken to be the 'truth' so that the assimilation update can be made perfectly. All the thermospheric and ionospheric fields are known, so there are none of the errors typically associated with the data assimilation process. The challenge is to produce accurate forecasts even when the external solar and geomagnetic drivers are significantly underestimated, as they are likely to be during a real storm. Such forecasts could be used for mission planning and early-warning systems. In this experiment, fields from the 'truth' run replace fields from a run with underestimated drivers to simulate an ideal assimilation scheme. As a first step towards creating a real assimilation scheme, the most successful simulated assimilation run is compared with real observations of vertical TEC. These observations are described in Section 7.1.2.

7.1.1 TIEGCM

The Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIEGCM) is a coupled model of the thermosphere-ionosphere system, described by *Richmond et al.* [1992]. TIEGCM is used in this experiment because it contains the physics necessary to represent the interactions and temporal development of the thermosphere-ionosphere system and because it has been validated and approved for public release. The model has been extended upwards to include the magnetosphere in the Magnetosphere Thermosphere Ionosphere Electrodynamics General Circulation Model (MTIEGCM) and downwards to include the mesosphere in TIMEGCM. These extended versions of the National Center for Atmospheric Research (NCAR) General Circulation Model have not yet been approved for public release. TIEGCM is the first in the line of NCAR general circulation models to include a calculation of electric fields and currents from the dynamo effects of the thermospheric winds. These fields and currents are included in the calculations of neutral and plasma dynamics. The original three-dimensional Thermospheric General Circulation Model (TGCM) is described by Dickenson et al. [1981]. TGCM had the same 5 degree grid that is used by TIEGCM, but TIEGCM has been extended to use 39 constant pressure surface layers as opposed to the 24 levels in the original

model. The thermosphere is principally driven by heating caused by the absorption of solar EUV and UV radiation, characterised in the model by the *F10.7* solar index, as well as by auroral heating. TIEGCM also takes in forcing from the geomagnetic *Kp* index and lower boundary processes, such as tides. *Roble et al.* [1988] coupled an ionospheric model to TGCM in order to create the Thermosphere-Ionosphere General Circulation Model (TIGCM). TIGCM includes a self-consistent description of the thermosphere-ionosphere system. The model calculates global distributions of *O*, O_2, N_2, *NO*, $N(^2D)$, $N(^4S)$, O^+, NO^+, O_2^+, N_2^+, N^+, electron density and ion temperature as well as neutral winds, temperature and major composition. The thermosphere and ionosphere are mutually coupled at every timestep. Electron density is calculated as the sum of the ion densities, so any changes made to the electron density model fields are lost almost immediately. Other three-dimensional coupled thermosphere-ionosphere models are available, including the Coupled Thermosphere Ionosphere Plasmasphere with electrodynamics (CTIPe) by *Fuller-Rowell et al.* [2002] and the Global Ionosphere-Thermosphere Model (GITM) by *Ridley et al.* [2006]. Incidentally, CTIPe and GIM both use the same electrodynamic formulation as TIEGCM.

7.1.2 Vertical TEC measurements

In order to gauge the potential of a real assimilation scheme using TIEGCM, the simulated assimilation results are compared with real observations of the ionosphere. The important vertical TEC parameter is used here. Vertical TEC is the sum of the free electrons in a vertical column. This is calculated from GPS observations and stored in global maps of vertical TEC, the production of which is described by *Rideout and Coster* [2006]. Vertical TEC can also be extracted from GPS-derived tomographic images of electron density. Ionospheric imaging techniques are described in detail in Section 3.5.

7.1.3 The storm of Halloween 2003

Between 29 October and 1 November 2003, a severe solar storm occurred that included periods of elevated geomagnetic activity caused by two Coronal Mass Ejections (CMEs). Solar and geomagnetic indices measured during this period, as well as a more typical control period, are used to drive TIEGCM in this study. Dramatic effects caused by this storm were observed in various aspects of the

thermosphere-ionosphere system. *Sutton et al.* [2005] found that density observations from CHAMP exhibited enhancements of 200-300% at altitudes of around 410 km during periods of maximum geomagnetic activity. In addition, the authors showed the accuracy of an empirical model, the Horizontal Wind Model 1993 (HWM-93), was compromised at low latitudes in disturbed conditions. HWM-93 overestimated the daytime westward wind intensification during severely disturbed conditions but underestimated westward wind speeds under moderately disturbed conditions. *Immel et al.* [2008] studied the same storm using a combination of CHAMP and Far Ultraviolet Spectrographic Imager (FUSI) observations. The authors attributed the large temperature and density changes observed by CHAMP at around 400 km to traveling atmospheric disturbances as well as to the direct effects of Joule heating. These large temperature and density changes were caused by impulsive auroral inputs as well as downwelling in the global thermospheric storm-time circulation. *Richmond and Lu* [2000] explained that storm-time heating causes a polar upwelling that results in equatorward flow in both hemispheres (winter and summer). This equatorward flow results in downwelling at mid-latitudes, which moves the air into regions of increasing pressure and therefore results in compressional heating. The result is that auroral heating is spread to all latitudes, leading to an expansion of the thermosphere that causes large increases in neutral density at high altitudes.

The effects of the Halloween 2003 storm on TEC in the American longitude sector are difficult to model due to two potentially competing effects. *Heelis et al.* [2009] showed that the expansion of a high-latitude electric field to mid-latitudes could potentially result in a day-side TEC enhancement of up to 300 TECU. However, *Sojka et al.* [2012] showed that storm-time thermospheric winds had a potentially equally large effect as the electric field changes. Those authors suggested both mechanisms could result in a 10-20% change in TEC and that the neutral winds could either reinforce or counteract the effects of the electric field enhancement. The storm of Halloween 2003 makes an ideal case study for this simulation experiment because it exhibits extreme external drivers and because it is recent enough to have been well-observed.

7.2 Method

The overall aim is to determine whether a data assimilation scheme can produce accurate ionospheric forecasts in the absence of good external driver estimates. In order to establish this, several experiments are performed. Each of these experiments follows the same general concept. First, TIEGCM is run for a storm period in order to create a 'truth' against which to compare forecasts. Then a geomagnetically quieter period, referred to here as 'typical', is selected from a few days before the storm. The storm run is repeated, but at a given point the model is switched over to use the drivers from the typical period. By comparing the results of this hybrid run with the 'truth' run, we can establish the forecast accuracy that could be achieved with a perfect assimilation update and a perfect model using incorrect drivers. This generic procedure is shown in Figure 22.

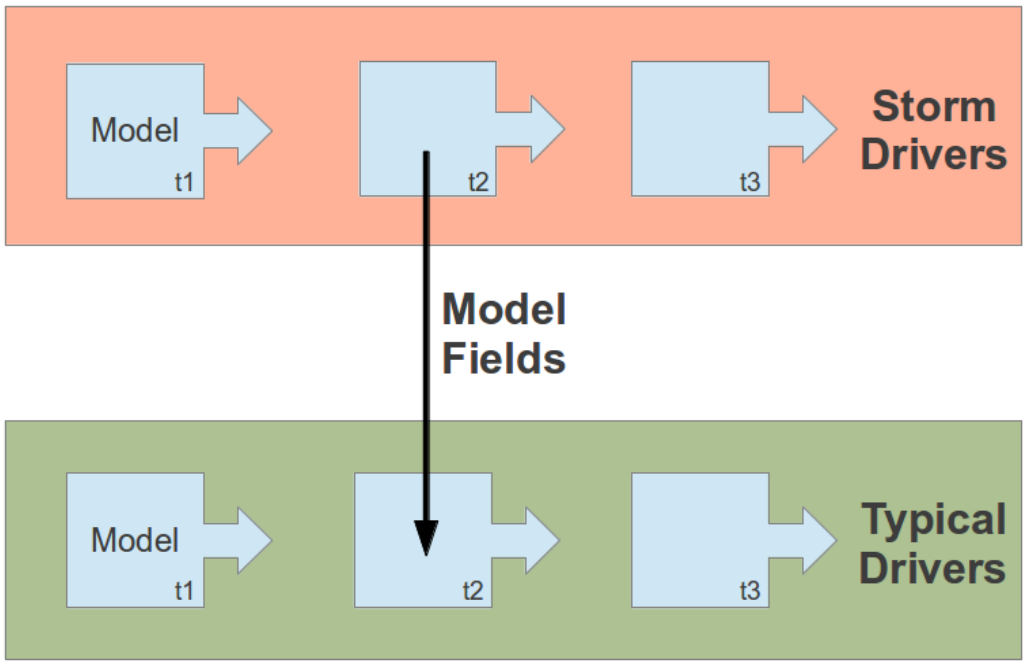

Figure 22: The general procedure followed in this chapter. Two model runs proceed from time t1 to t3 under the influence of different drivers. At time t2, fields from the storm-time run are put into the typical run. This run then proceeds to time t3 under the influence of the typical drivers. The drivers are Kp, F10.7, IMF B_x, B_y and B_z, cross-tail potential and hemispheric power. The model fields are global arrays of temperature, wind, density and composition.

In the procedure described above, the drivers from the typical period are used to represent the drivers that could have been forecast for the storm – the values are underestimated compared with the real storm values, but they are typical of the period during which the storm occurs. The procedure simulates a perfect assimilation scheme because the forecast model's initial state is identical to the initial 'truth' state. The procedure also simulates the use of a perfect model because the same model is used for both runs.

There may not be sufficient observations to constrain the whole system, so a data assimilation scheme might have to specify only some of the model fields. It would be useful to identify which groups of model fields have the greatest effect on forecast accuracy so that future observation campaigns can be directed towards measuring those fields. The effects of changing different groups of model fields on forecast accuracy can be tested by combining some fields from the 'truth' run with other fields from the typical period in order to create a hybrid model state. This presents a greater risk than replacing the full set of fields because the resulting hybrid state might be unstable. However, we found no evidence of such instabilities in our experimental results. In this chapter, different groups of model fields in the typical period run will be replaced with fields from the storm-time run. The hybrid run is then continued under the influence of the typical drivers. This is done in order to determine which model fields are most important to the progression of the model state over time. An assimilation scheme could then focus on specifying the important fields more accurately in order to improve forecasts. In addition, these experiments should establish the maximum forecast accuracy that is achievable with underestimated drivers.

The experiment described in Section 7.3.1 determines whether the initial thermospheric or ionospheric model fields are more important to plasma density forecasts, using the storm of Halloween 2003 as a case study. The model is run for 29 and 30 October to cover the storm and for 15 and 16 October to represent a typical period. This typical period run is then repeated three times. Each time, some model fields are replaced with fields from the storm-time run at 6 am on 15 October. The different groups of fields replaced are listed in Table 3 and identified as: (a) – ionospheric fields, (b) – thermospheric fields and (c) – all fields. After the replacement step, each model run is continued using the typical day drivers.

Run type	Replacement fields
(a) Ionosphere	Ne, O^+, O_2^+, Ti, Te
(b) Thermosphere	Tn, Un, Vn, O, O_2, NO, N(4S), N(2D)
(c) Full state	Ne, O^+, O_2^+, Ti, Te, Tn, Vn, Un, O, O_2, NO, N(4S), N(2D)
(u) Typical period (unchanged)	none

Table 3: Model fields replaced in the experiments described in Sections 3.1 and 3.2. The first two runs are intended to determine whether ionospheric or thermospheric initial conditions are more important to the forecast plasma density. The third run is included to control for synergistic coupling effects and to establish an upper limit for the effectiveness of full data assimilation. The final run is included to establish a lower limit for forecast accuracy. Where applicable, the fields include the previous-timestep version of the listed variables.

The experiment described in Section 7.3.2 validates the results shown in 3.1 by repeating that experiment with the storm of 20 and 21 November 2003. Once again, the replacement fields are identified in Table 3. This time, the typical period is 16 – 17 November 2003.

The experiment described in Section 7.3.3 determines which of the thermospheric fields (neutral winds, neutral temperature and neutral composition) is most important to plasma density forecasts, once again using the storm of Halloween 2003. The procedure of Section 7.3.1 is repeated using different groups of replacement fields. These are listed in Table Error: Reference source not found and identified as (d) – neutral winds, (e) – neutral temperature, (f) – neutral composition and (g) – full thermosphere.

Run type	Replacement fields
(d) Neutral winds	Un, Vn
(e) Neutral temperature	Tn
(f) Neutral composition	O, O_2, NO, N(4S), N(2D)
(g) Full thermosphere	Tn, Vn, Un, O, O_2, NO, N(4S), N(2D)
(u) Typical period (unchanged)	none

Table 4: Model fields replaced in the experiments described in Section 7.3.3. These runs are intended to determine which aspect of the thermospheric initial conditions is most important to the forecast plasma density. The full thermosphere run controls for synergistic effects. The final run is included to establish a lower limit for forecast accuracy. Where applicable, the fields include the previous-timestep version of the listed variables.

The *Kp* values observed on 15 and 29 October 2003 are shown in Figure 23 in order to illustrate why 15 October 2003 was chosen as a typical day and 29 October 2003 was chosen as a storm-time case. The 15th had moderate solar and geomagnetic activity for that period, whereas the 29th had some of the highest driver values of the Halloween 2003 storm. *F10.7* was 275 on the storm day and 95 on the typical day. This is a very large difference since *F10.7* generally varies between about 70 and 300. *Kp* values started off similar on both days, but storm values hit the maximum possible value of 9 at 06:00 UT. For that reason, the replaced-field runs were started at 06:00 UT. This way, *Kp* was consistently much higher in the storm-time run throughout the test period.

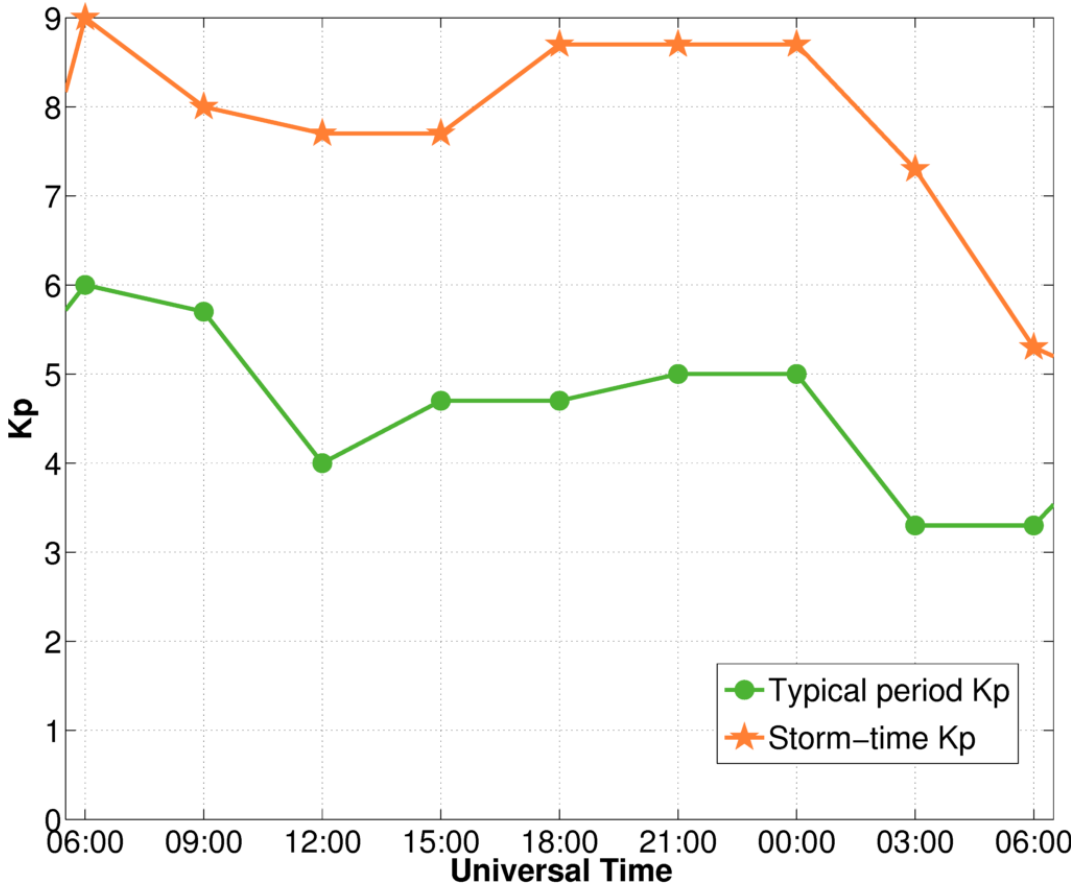

Figure 23: Kp on the storm day (29 October 2003) in orange and on the typical day (15 October 2003) in green.

7.3 Results

7.3.1 Model runs with replaced fields

This experiment is designed to measure the effects of changes to the initial ionospheric and thermospheric model fields on electron density forecast accuracy. Three groups of replacement fields are chosen: one to represent ionospheric assimilation, one to represent thermospheric assimilation and one to represent full state (ionosphere plus thermosphere) assimilation. The model fields that were replaced are described in Table 3. If there are no instabilities caused by the creation of a hybrid thermosphere-ionosphere system, the combined effects of the ionospheric and thermospheric replacement runs should match the effects of the full state replacement run.

The global RMS electron density error of the replaced model field runs is shown in Figure 24. The 'truth' here is the 29-30 October 2003 storm run and the typical period

is the 15-16 October 2003. Predictably, each modified run underestimates the electron densities because the external drivers are not able to sustain the production rates necessary to match the storm conditions. From this graph, it can be seen that each replaced-field model run heads back towards the typical day state. However, changes to the ionosphere and thermosphere initially improve the forecasts.

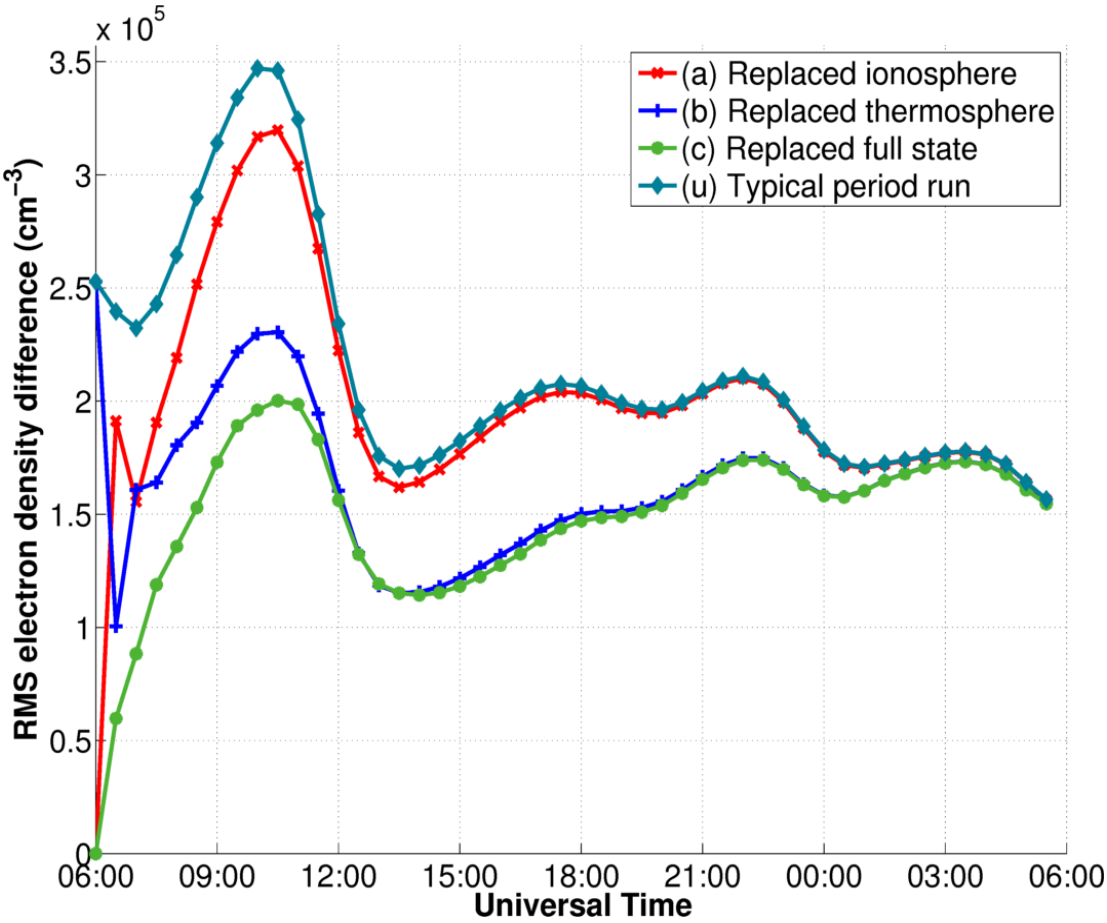

Figure 24: The global RMS electron density difference of four runs with the 'truth' model run starting on 29 October 2003. Each of these runs used the drivers from 15 October 2003, but had some model fields replaced with ones from 29 October 2003.

Changes to the ionosphere (a) initially had more impact than changes to the thermosphere (b) but 90% of the electron density forecast improvement of run (a) was lost through recombination after 3 hours and 30 minutes. In other words, run (a) relaxed back to within 10% of the typical day run, (u), by 09:30 UT. The replaced ionosphere run, (a), relaxed back to the typical day run, (u), which had no replaced fields. The ionosphere's nightly return to low levels of ionization explains why any changes are normally wiped out within this timeframe. One mechanism for ionospheric changes to persist longer than 12 hours would be for the changes to feed

into the thermosphere and then back to the ionosphere. It appears this feedback effect did not occur here, even with a very strong update. The changes to the thermosphere, shown in (b), resulted in a persistent improvement in agreement with the storm-time run that lasted for the duration of the test period. The thermospheric run, (b), returned to within 10% of the typical period run, (u), after 18 hours and 30 minutes. This is not surprising since the thermosphere is less strongly driven than the ionosphere (especially if we consider the thermosphere as one of the ionospheric drivers and vice-versa). It is natural that the initial values of the system will continue to affect its behaviour for much longer, and that thermospheric changes will feed into improved ionospheric specification. The ionospheric change, (a), only produced better agreement than the thermospheric change, (b), at the first timestep, before the thermospheric change had any chance to affect the ionosphere. As a side note, the improvement in forecast accuracy of the run starting with all storm-time fields shown in (c), was generally about equal to the sum of the improvements from (a) and (b). This shows that the model was not unbalanced by starting from artificial hybrid states such as (a) and (b) and that any coupling effects were of secondary importance. The full-state run, (c), also returned to within 10% of the typical period run after 18 hours and 30 minutes.

The run using the full storm-time initial conditions with drivers from the typical day, (c), produced the best forecast of the storm. This shows that accurate specification of initial thermosphere-ionosphere conditions is an important part of the forecasting process. Even without accurate driver specification, data assimilation has the potential to improve forecasts for over 18 hours. In practice, plasma density is the only well-observed model field in the system, so it might be impossible to specify the other model fields accurately. However, *Matsuo and Araujo-Pradere* [2011] demonstrated, using an ensemble Kalman filter Observation System Simulation Experiment (OSSE), that inclusion of thermospheric variables in the assimilation state vector reduced electron density forecast error. The experiment assimilated simulated observations of electron density, so all changes to the thermosphere depended on the cross-covariances calculated from the ensemble spread. A successful assimilation scheme will need to use correlations between plasma density and other fields to specify the important parts of the system. From these results, it appears that thermospheric model fields should not be omitted from an assimilation scheme if the intention is to forecast for more than a few hours. The

experiment in Section 7.3.3 will determine which of the thermospheric model fields is most important to the storm-time ionospheric electron density forecast. In Section 7.3.2, the results shown in this section will be validated by repeating the above experiment with a different storm.

If there are no instabilities or coupling effects caused by the creation of a hybrid thermosphere-ionosphere system, the combined effects of the ionospheric and thermospheric replacement runs on the ionospheric plasma density forecast accuracy should match the effects of replacing the full state. In this case, the combined forecast accuracy improvement from the ionospheric and thermospheric replacement runs matched the forecast accuracy improvement of the full state replacement to within 20% of the full state improvement at all times and to a mean of 2.5% over the test period. These results show that, whilst there are some coupling effects, those effects generally do not have a significant effect on the forecast accuracy of different runs.

7.3.2 Validation of Halloween storm results

Although the results of the experiment described in Section 7.3.1 appear to be consistent with theory, it could be that unidentified physical processes have conspired to produce a seemingly reasonable set of results. The storm of Halloween 2003 did see a double CME that is not common in solar storms. In general, these solar storms induce variability in the Earth's magnetic field, which is measured in the Kp values that are used to drive TIEGCM. In this section, the experiment from Section 7.3.1 is repeated using the storm of 20 November 2003. Although this storm was not as spectacular as that of Halloween 2003, *Kp* still reached 9^- at 15:00 UT and 18:00 UT while *F10.7* was 171 that day. A typical day was identified as 16 November 2003. This day was used for comparison with the storm-time results. Once again, runs were started at 06:00 UT through to 06:00 UT the next day so as to simulate an assimilation just before the spike in geomagnetic activity. This typical period had *Kp* between 4 and 5^+ and *F10.7* at 102.

The results shown in Figure 25 support the conclusions drawn in Section 7.3.1. Once again the run starting with all storm-time fields, (c), has the lowest error, remaining substantially more accurate for over 12 hours after the simulated assimilation, while

the run with replaced thermospheric fields (b) performs much better than the run with replaced ionospheric fields (a) for the same period. The run with replaced ionospheric fields, (a), relaxes back to within 10% of the typical period run, (u), within four hours. The run with replaced thermospheric fields, (b), relaxes back to within 10% of the typical period run, (u), within 15 hours and 30 minutes, while the replaced full state run, (c), takes 16 hours. This provides an upper bound on the potential benefits of assimilation for forecasting. The relaxation times are somewhat shorter than in Section 3.1 because the storm chosen for study in this section is less intense.

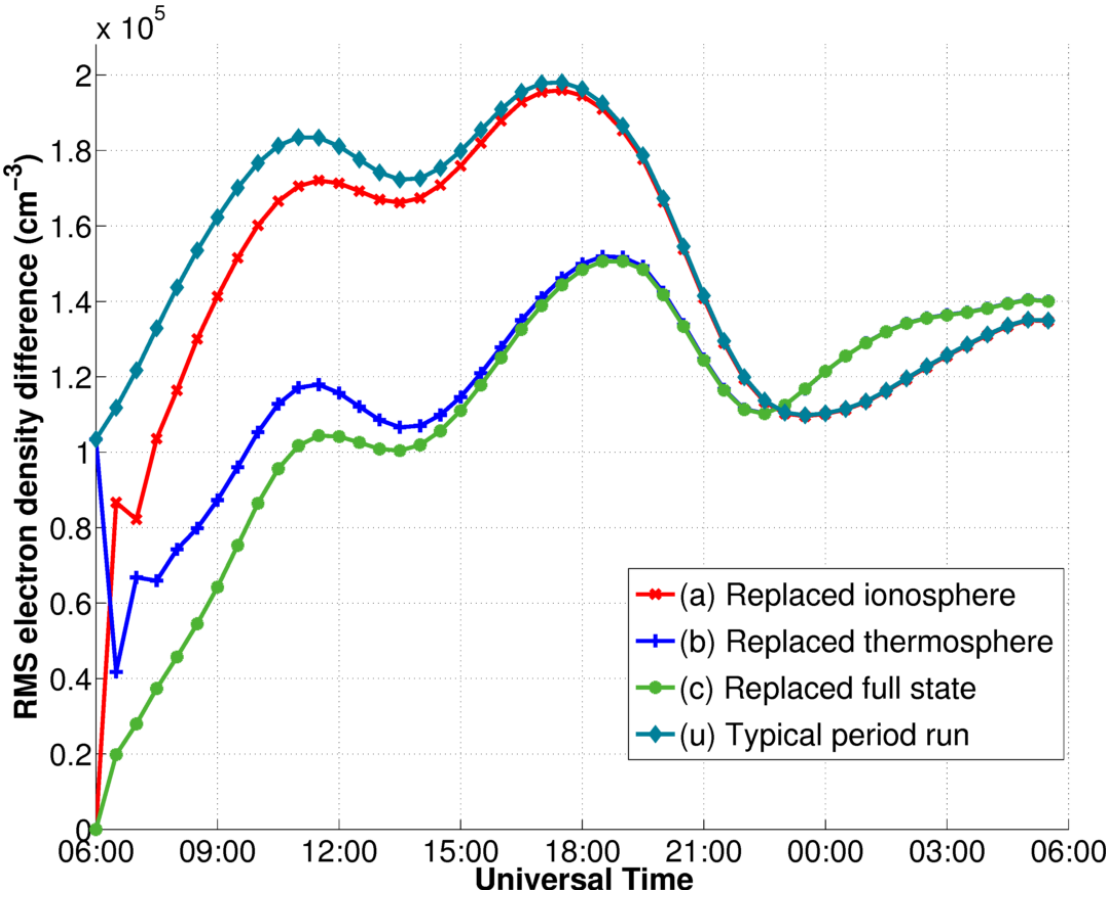

Figure 25: The global RMS electron density difference of four runs with the 'truth' model run starting on 20 November 2003. Each used drivers from 16 November 2003, but had some initial fields from 20 November 2003.

7.3.3 Model runs with replaced thermospheric fields

While the accurate initial specification of all model fields is clearly desirable in terms of forecasting, in reality some compromises have to be made. Direct observation of all variables to the extent that would be required to specify the full thermosphere-

ionosphere state would be prohibitively expensive. By identifying those variables most important to ionospheric forecasting, it will be possible to target future observation campaigns towards specifying the most valuable model fields. Groups involved in the development of space- and ground-based instruments may wish to target their efforts accordingly. In addition, known relationships with other, better-observed variables might be used to update the most important model fields. This will be most effective if just the important variables are included in the assimilation state vector.

In order to determine which thermospheric fields are most important to the ionospheric electron density specification, each different thermospheric field is replaced in separate runs. With reference to Table 4, the different thermospheric fields to be replaced are the neutral winds (d), the neutral temperatures (e) (both zonal and meridional) and the neutral composition (f). The run with all thermospheric fields replaced (g) shows the combined effect of making all the changes in runs (d), (e) and (f), while the model run for 15 October 2003 with no fields replaced, (u), is included as a baseline that shows the forecast quality to be expected with poorly specified initial conditions and external drivers.

The results in Figure 26 show that neutral composition is by far the most important model field to specify accurately if the intention is to improve ionospheric electron density forecasts. In fact, the run with the storm-time neutral composition field (f) outperforms the run with the full set of thermospheric fields (g) for a brief period during the first three hours. The storm-time neutral composition run, (f), takes 19 hours and 30 minutes to return to within 10% of the typical period run, (u). This is an hour longer than the full thermosphere run, (g). The other runs – replaced neutral winds (d) and temperatures (e) – have a negligible impact, with both runs under-performing the run with no replaced fields, (u), at times. Neutral composition affects ionization and recombination rates, so it follows that modifications to this model field should result in improved electron density forecasts. It is worth considering the various constituents that make up neutral composition, all of which are listed in Table Error: Reference source not found. Of these, the ratio of O and N_2 is likely to have the most important effect on storm-time ionospheric plasma densities. This is discussed in more detail in Section 7.4.

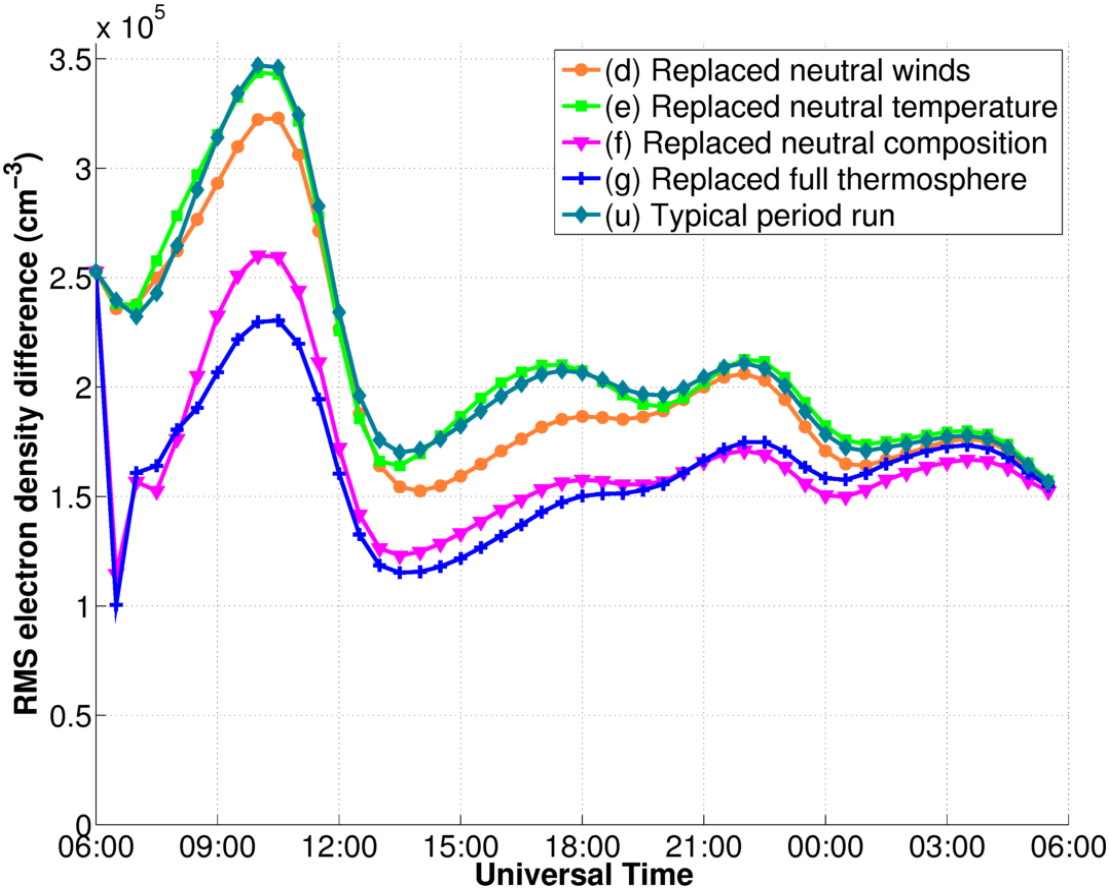

Figure 26: The global RMS electron density difference of five runs with the 'truth' model run starting on 29 October 2003. Each of these runs used the drivers from 15 October 2003, but had some thermospheric model fields replaced with ones from 29 October 2003.

7.3.4 Comparison with storm-time GPS TEC observations

The results described above show that a well-specified initial assimilation update could result in model forecast improvement compared with the model run that uses the correct drivers and initial conditions, which is taken to be correct. The results described in this section are intended to test the model's performance against real data, in order to establish how accurately the model can perform under optimal conditions. Once again, the Halloween 2003 storm is used as the test period with typical drivers coming from 15 – 16 October 2003. The storm-time model run, used as the 'truth' against which to compare the runs in Figure 26, is compared with globally distributed vertical TEC observations calculated by a slant-to-vertical correction. Results of the run with replaced thermospheric composition, (f), are included to show what forecast improvements could be achieved if thermospheric

composition was defined more accurately. The typical period run, (u), is also included to show how much difference the replacement of the thermospheric composition has made.

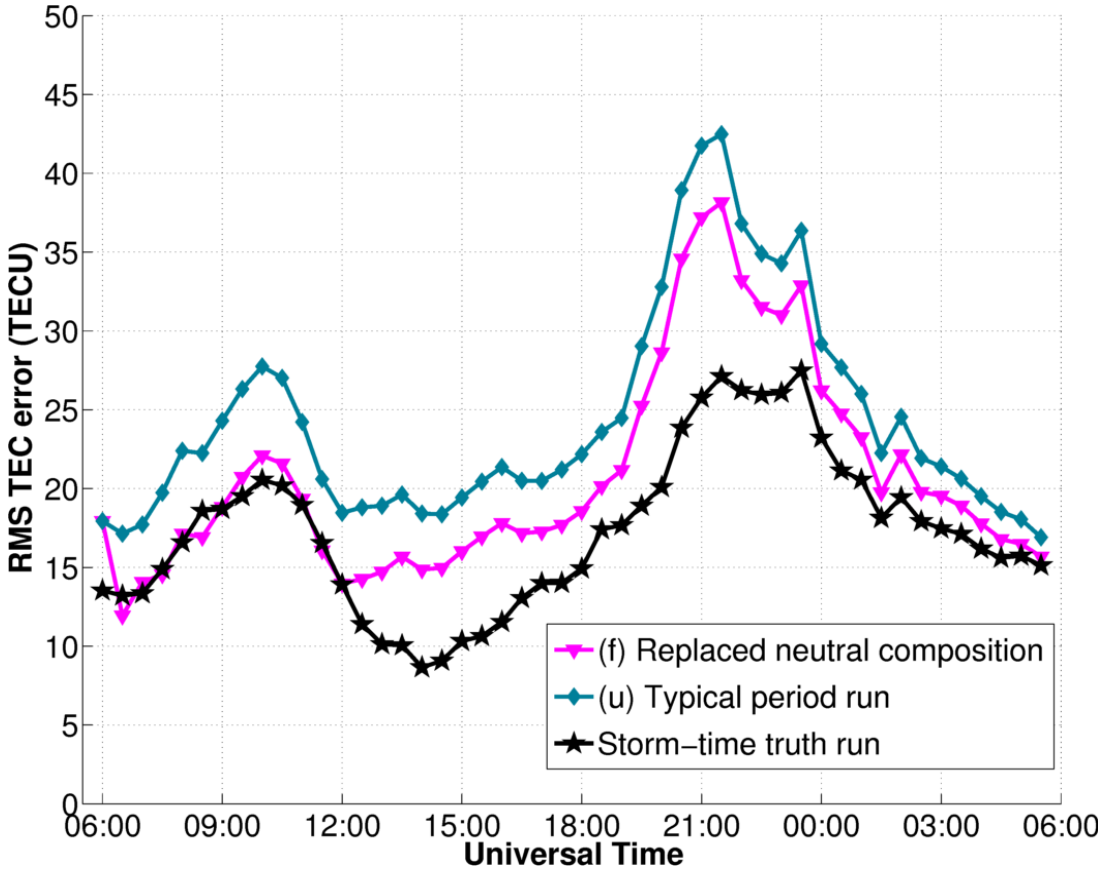

Figure 27: RMS TEC error of three TIEGCM model runs compared with GPS observations of the real vertical TEC. Storm-time run (previously used as the 'truth'), with correct initial conditions and external drivers, is shown in black. The typical day run, (u), with incorrect initial conditions and external drivers, is shown in light blue. Results run from 06:00 UT on 29 October 2003 to 06:00 UT on 30 October 2003.

The results in Figure 27 show that, over the selected period, the model performed best when provided with the correct initial conditions and external drivers. Starting with the correct neutral composition but the other initial conditions and the drivers from the typical day, as in run (f), the model was several TECU more accurate than the typical day run in reproducing observed TEC. Somewhat surprisingly, this improvement persisted for the whole 24 hour period. The results show that the model is inaccurate even with the correct drivers and initial conditions, but changes to the initial conditions can produce long-lasting forecast improvements.

7.4 Discussion

The results show that thermosphere-ionosphere data assimilation could produce significant improvement to forecasts of ionospheric electron density and TEC under optimal conditions. This is supported by the results shown in Figures 24 and 25, where a full thermosphere-ionosphere assimilation improved forecast accuracy by at least 10% for 18.5 hours and 16 hours respectively. In situations with both an accurate model and a good update of the full model state but without good external driver forecasts, assimilation can produce up to a 50% improvement in agreement with the 'true' state compared to a model run without the update or the correct drivers. However ionosphere-only updates, such as are shown in run (a) in Figures 24 and 25, give less than 10% improvement after four hours. These results support the findings of *Jee et al.* [2007], who showed that an ionospheric update of the Thermosphere-Ionosphere Nested Grid (TING) with values from the USU GAIM scheme resulted in an *e*-folding time (the time in which the difference between the electron densities in the assimilation and control runs reduced by a factor of *e*) of about 2-3 hours under most conditions. The authors expected to find a different result if they had included an update of thermospheric fields in addition to the ionospheric update.

The results in Figure 27 show that the storm-time truth run reproduces the observed TEC most accurately during the Halloween 2003 storm. This is to be expected as that run uses all the correct initial model fields and the correct external drivers. However, run (f), which uses the initial storm-time neutral composition but all other model fields and external drivers from the typical period run is almost as good for the first six hours. This is interesting as it suggests that accurate specification of the neutral composition may compete with the accuracy of external driver specification for short-term plasma density forecasting. It is important to note that this situation is reversed after six hours. Given that forecasts of external drivers may not be available or accurate in a storm-time forecasting situation, assimilation of the most important model fields could could be the most viable way to improve forecasts. None of the runs shown in Figure 27 were especially accurate, which shows that modelling improvements are also essential for accurate forecasts. The current crude driver specification is also a source of error – it should be possible to improve accuracy by

using more informative measurements than *Kp* and *F10.7*. For example, the full observed spectrum of solar flux could be used directly, rather than relying on a daily average spectrum generated solely from observations of 10.7 cm flux. A data assimilation scheme would have the advantage of influencing the model with real observations, so it should be possible to get more accurate results than those shown here.

This study addresses the issue of forecasting storm-time ionospheric plasma densities with only the information available at the beginning of the storm. Quite different results might be expected if the exercise was repeated from the middle of the storm – it is likely that external driver forecasts for the remainder of the storm would be more accurate at that point, for example. The forecast time used in this chapter was chosen to be close to the beginning of the storm. This time was chosen because it is more important to forecast a storm before it happens than it is to forecast the after-effects of a storm whose major effects have already been observed. However, the authors acknowledge that there are applications for which a variety of forecasting scenarios are relevant. These situations could be explored in future studies.

Systematic biases in the model may be overcome with an assimilation scheme. An ensemble Kalman filter approach to data assimilation allows for the tuning of certain parameters to correct model biases on an ongoing basis. These could be parameters such as the Burnside factor (a scaling factor of the O+-O collision frequency) . Once identified, these parameters can be included in the state vector and updated at each assimilation step. A future study should identify those parameters that are important to the evolution of forecasts and are poorly specified. The presence of observations across the globe is required to prevent the scheme from tuning the entire model state to correct a local bias. The bias does not need to be permanent because the variable can be continually updated, but the bias must be long-lasting enough to be observed at the assimilation update time and still be in effect for the duration of the forecast.

The results in Figure 26 show that the neutral composition is the most important thermospheric field to update in order to improve plasma density forecasts. The ratio of O to N_2 is recognised to be an important component of storm-time neutral composition, since modelling work by *Rishbeth and Müller-Wodarg* [1999] indicated

that the O to N_2 ratio is altered by high-latitude energy inputs and affects peak plasma densities, *NmF2*. This raises the question of how to specify the ratio of O to N_2 accurately. The most straightforward approach would be to assimilate observations directly – for example with measurements from the Special Sensor Ultraviolet Spectrographic Imager [*Paxton et al.,* 1992] – but this approach is limited by data availability. Alternatively, it might be that changes to the model's neutral composition could be achieved indirectly, by repeated assimilation of other model fields such as neutral temperature. Once again this approach is limited by data availability, but the approach may also suffer from an additional problem – changes to other model fields are likely to take longer to feed into thermospheric composition changes than direct observations of the thermospheric composition. A third potential approach is to make use of known correlations between model fields. This would mean using observations of electron density and other well-observed parameters to adjust the neutral composition. Clearly there are risks attached to this approach. Further work should investigate whether it is possible to improve estimates of thermospheric composition by using observations of other fields.

The results in Figure 27 show that the TIEGCM model has significant inaccuracies, with RMS errors between 8.6 and 27.5 TECU during the test period. These errors would be best dealt with by improvements to the model itself, although accuracy could also be improved by characterizing the important drivers in a more complete and detailed manner. For example, direct observations of the full spectrum of solar flux could be used instead of assuming a standard spectrum based on the observed *F10.7* values. There are ongoing efforts to improve TIEGCM, notably the extension down to the mesosphere in TIMEGCM and up to the magnetosphere in MTIEGCM, but it must be recognised that accurate modelling of the thermosphere-ionosphere system remains a significant challenge. It should be possible to reduce the model's random errors and systematic biases in short term forecasts by using data assimilation. The results presented here show that a model run with one set of drivers can be made to behave more like a model run with another set of external drivers for over 12 hours by making changes to the initial conditions. Future work should investigate whether a model that assimilates real ionospheric observations can produce improved plasma density forecasts.

In summary, the results show that thermospheric data assimilation could provide

plasma density forecast improvements of at least 10% for over 18 hours, provided that it is possible to specify the neutral composition accurately. However, 90% of the effects of ionospheric data assimilation are lost within four hours.

With the exception of the vertical TEC observations, which are available from MIT Haystack's Madrigal service, all the results presented in this chapter are available online at www.bath.ac.uk/elec-eng/invert/tiegcm_results.

8 Conclusions and further work

At the start of this research, several problems existed in ionospheric specification. The results presented in Chapter 4 established the current performance of ionospheric specification techniques, including ionospheric tomography and empirical modelling. Chapters 5 and 6 addressed imaging problems caused by a lack of observations in the horizontal and in the vertical. Improvements were made by adding observations and by using known physics to constrain the problem. In Chapter 7, the problem of forecasting a storm was addressed with reference to the physics of the combined thermosphere-ionosphere system. These results provide a foundation for the development of an ionospheric forecast system.

The results of Chapter 4 demonstrate that ionospheric tomography and empirical modelling can both specify monthly median ionospheric TEC over all solar cycle conditions in regions of good GPS coverage. Empirical modelling produces errors in situations where the monthly averaged observed input parameters are significantly inaccurate, which can happen during periods of intense solar activity. Global GPS coverage is sufficient to produce accurate low resolution ionospheric images that show known ionospheric phenomena, such as the annual asymmetry. The degree of agreement between the IRI empirical model and the MIDAS tomography algorithm suggests the two approaches could be combined successfully to bridge data gaps and provide increased resolution in areas of good observation coverage. The extensive series of ionospheric images produced and validated in this study provide a resource that could be used to validate and benchmark future approaches to the problem of ionospheric specification. New approaches should reproduce the bulk features shown in this data set whilst offering some tangible benefit over the current technique – either by improving resolution, by dealing with gaps in observation coverage or by operating in real time or forecasting modes.

In Chapter 5, the problem of horizontal resolution was addressed with reference to ionospheric imaging in Africa. A simulation approach demonstrated that the current operational GPS receivers available in Africa do not provide sufficient observations

to capture the expected horizontal ionospheric TEC structures. An extended campaign network provided far more accurate images. Remaining inaccuracies were shown to be linked to the type of structures present, as well as to the level of observation coverage. A fictional, ideal network of GPS receivers produced the best simulated images, as expected, but did not remove errors entirely. This was because the addition of ground-based GPS receivers to an already dense network did not significantly add to the range of ray-path angles, or the number of rays available. Image quality would be more efficiently improved by the addition of other types of ionospheric observation, or by the extension of existing GPS receivers to use other GNSS systems. A future study should explore the relationship between ionospheric image accuracy and GPS positioning accuracy in Africa. There is a direct relationship between the line-of-sight signal delay and the slant TEC along a satellite-to-receiver path, but the positioning errors experienced by users depend on the number and location of available satellites as well as the TEC present. *Allain and Mitchell* [2009] presented a method for calculating positioning error from maps of electron density that could be applied to the maps presented in Chapter 5. Future work should also seek to translate these scientific results into metrics that are directly useful for other applications.

The lack of vertical resolution in ionospheric images is a significant problem for users interested in applications such as HF ray-tracing as well as those looking to understand the physics of the ionosphere. In Chapter 6, a new method for improving the vertical resolution of ionospheric images was presented. The new approach involves including ionosonde observations of the vertical electron density profile into GPS tomography. Using the Jicamarca Incoherent Scatter Radar as a ground truth, it was shown that the vertical profile of ionospheric images can be greatly improved by including ionosonde observations. This approach is robust enough to work with auto-scaled ionosonde observations, which are of poorer quality than manually scaled ionograms but are available in near-real-time. A group with suitable operational capabilities should develop this technique into a near-real-time ionospheric specification system.

Given the conclusion of Chapter 4, that the IRI 2007 empirical model generally agrees well with MIDAS images of the ionosphere, it should be possible to extend

the technique presented in Chapter 6 to take advantage of observations from multiple ionosondes in a GNSS-based ionospheric imaging algorithm simultaneously. The approach presented in Chapter 6 is limited to using only a single ionosonde because it uses the same set of vertical basis functions at each latitude/longitude grid-point. This approach is not suitable for accommodating diverse observations from multiple ionosondes at different geographic locations because a single set of basis functions will not be able to reproduce the different profiles from each ionosonde. A more sophisticated approach would be to use an empirical model to ingest ionosonde observations at multiple locations and interpolate between them. The resulting ionospheric specification could be used to create a set of spatially varying basis functions that could accommodate observations from multiple ionosondes as well as from ground- and space-based GNSS.

Ionospheric forecasting is a challenging extension to the problem of ionospheric specification. A persistence forecast will not produce good results during disturbed conditions, so some model of ionospheric dynamics is necessary for storm-time forecasting. If a global forecast algorithm was implemented, the model could also carry information from well-observed regions to areas of poorer coverage. The study presented in Chapter 7 showed that it is far more important to specify the thermosphere than the ionosphere if the intention is to forecast storm-time plasma densities. This finding highlights the need for a physics-based model of the thermosphere in a successful ionospheric forecast system. Without a thermospheric model, there is no way of carrying information from thermospheric observations forwards in time. Effort should be devoted to the creation of a combined thermosphere-ionosphere data assimilation approach. In the long term, such an approach could be extended by detailed coupling to lower and middle atmosphere models and by improved specification of the important solar and geomagnetic drivers. Such an approach would offer the additional benefit of improving the thermospheric specification, which is useful for satellite drag calculations.

The study presented in Chapter 7 showed that current ionospheric models do not accurately reproduce storms. Future work should investigate whether ionospheric models are currently limited by their internal physics and approximations, or by the crude nature of external driver indices such as *Kp* and *F10.7*. Real data assimilation

algorithms will face a second problem that was set aside in Chapter 7 - lack of observations. The findings of that study should serve as motivation for observationalists to provide global, near-real-time measurements of thermospheric composition, and for data assimilation groups to make use of the observations that are already available.

9 References

Abdullah, M., H. J. Strangeways & S. S. N. Zulkifli (2010), Ionospheric differential error determination using ray tracing for a short baseline, *Adv. Spa. Res.*, 46 (10), pp 1326-1333, doi:10.1016/j.asr.2010.07.005.

Allain, D. J., & C. N. Mitchell (2009), Ionospheric delay corrections for single-frequency GPS receivers over Europe using tomographic mapping. *GPS solutions*, 13(2), 141-151.

Andreeva, E. S., (1990), Radio tomographic reconstruction of ionization dip in the plasma near the Earth, *J. Exp. Theor. Phys. Lett.*, 52, 142 – 148.

Angling, M. J., & N. K. Jackson-Booth (2011), A short note on the assimilation of collocated and concurrent GPS and ionosonde data into the Electron Density Assimilative Model, *Radio Sci.*, 46, RS0D13, doi:10.1029/2010RS004566.

Appleton, E. V., & M. A. F. Barnett, (1925), On some direct evidence for downward atmospheric reflection of electric rays, *Proc. Royal Soc. London*. Series A, 109(752), 621-641.

Austen, J. R., S. J. Franke, C. H. Liu & K. C. Yeh (1986), Application of computerized tomography techniques to ionospheric research. In International Beacon Satellite Symposium on Radio Beacon Contribution to the Study of Ionization and Dynamics of the Ionosphere and to Corrections to Geodesy and Technical Workshop (Vol. 1, pp. 25-35).

Austen, J. R., S. J. Franke & C. H. Liu (1988), Ionospheric imaging using computerized tomography, *Radio Sci.*, 23, 299 – 307.

Azzarone, A., C. Bianchi, M. Pezzopane, M. Pietrella, C. Scotto & A. Settimi (2012), IONORT: A Windows software tool to calcolate the HF ray tracing in the ionosphere, *Comp. & Geosci.*, 42, doi:10.1016/j.cageo.2012.02.008.

Bailey, G. J., N. Balan, & Y. Z. Su (1997), The Sheffield University plasmasphere ionosphere model– A review, *J. Atmos. Sol. Terr. Phys.*, 59, 1541 – 1552.

Bernhardt, P. A., R. P. McCoy, K. F. Dymond, J. Ml Picone, R. R. Meier, F. Kamalabadi, D. M. Cotton et al. (1998), Two-dimensional mapping of the plasma density in the upper atmosphere with computerized ionospheric tomography (CIT), *Phys. Plasmas*, 5, 2010 – 2021.

Berryman, J.G., (2000), Analysis of Approximate Inverses in Tomography I. Resolution analysis of common inverses, *Optimization and Engineering*, 1 (1), pp 87-115

Bibl, K., & B. W. Reinisch (1978), The universal digital ionosonde, *Radio Sci.*, 13(3), 19–530, doi:10.1029/RS013i003p00519.

Bilitza, D., (2001), International Reference Ionosphere 2000, *Radio Sci.*, 36, 261 – 275.

Bilitza, D., & B. W. Reinisch (2008), International Reference Ionosphere 2007: Improvements and new parameters, *J. Adv. Space Res.*, 42(4), 599-609.

Bilitza, D., L. A. McKinnell, B. W. Reinisch & T. Fuller-Rowell, (2011). The international reference ionosphere today and in the future. *J. Geodesy*, 85(12), 909-920.

Blewitt, G., (1990), An automatic editing algorithm for GPS data, *Geophys. Res. Lett.*, 17(3), 199-202

Bracewell, R. N., (1956), Strip integration in radio astronomy, *Aust. J. Phys.*, 9, 198 – 217.

Breit, G., & M. A. Tuve (1925), A radio method of estimating the height of the conducting layer. *Nature*, 116, 357.

Bust, G. S., D. Coco, & J. J. Makela (2000), Combined ionospheric campaign: 1. Ionospheric tomography and GPS total electron count (TEC) depletions, *Geophys.*

Res. Lett., 27, 2849 – 2852.

Bust, G. S., T. W. Garner, and T. L. Gaussiran II (2004), Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm, *J. Geophys. Res.*, 109, A11312, doi:10.1029/2003JA010234.

Bust, G. S., & C. N. Mitchell (2008), History, current state, and future directions of ionospheric imaging, *Rev. Geophys.*, 46, RG1003, doi:10.1029/2006RG000212.

Cannon, P. S., G. Crowley, B. W. Reinisch & J. Buchau (1992), Digisonde Measurements of Polar Cap Convection for Northward Interplanetary Magnetic Field, *J. Geophys. Res.*, 97(A11), 16,877–16,885, doi:10.1029/92JA01190.

Censor, Y., (1983), Finite series expansion methods, *Proc. IEEE*, 71, 409 – 419.

Chapman, S., (1931), The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, *Proc. Phys. Soc. (London)*, pp 43, 26–45.

Chartier, A. T., C. N. Mitchell & D. R. Jackson (2012a), A 12 year comparison of MIDAS and IRI 2007 ionospheric Total Electron Content, *J. Adv. Spa. Res.*, *49*(9), 1348-1355.

Chartier, A. T., N. D. Smith, C. N. Mitchell, D. R. Jackson & P. J. Condor (2012b), The use of ionosondes in GPS ionospheric tomography at low latitudes, *J. G. R.: Spa. Phys. (1978–2012)*, *117*(A10).

Chartier, A. T., J. Kinrade, C. N. Mitchell, J. A. R. Rose, D. R. Jackson, P. J. Cilliers, J.-B. Habarulema et al. (2013a), Ionospheric Imaging in Africa, *Radio Sci.* doi:10.1002/2013RS005238

Chartier, A. T., D. R. Jackson & C. N. Mitchell (2013b), A comparison of the effects of initializing different thermosphere-ionosphere model fields on storm time plasma density forecasts, *J. Geophys. Res. Space Physics,* 118, doi:10.1002/2013JA019034.

Choi, B. K., J. H. Cho & S. J. Lee (2011), Estimation and analysis of GPS receiver differential code biases using KGN in Korean Peninsula, *J. Adv. Space Res.*, 47 (9), 1590-1599.

Ciraolo, L., & P. Spalla (1997), Comparison of ionospheric total electron content from the Navy Navigation Satellite System and the GPS, *Radio Sci.*, 32(3), 1071–1080, doi:10.1029/97RS00425.

Ciraolo, L., F. Azpilicueta, C. Brunini, A. Meza & S. M. Radicella (2007), Calibration errors on experimental slant total electron content (TEC) determined with GPS, *J. Geod.*, 81(2), 111 – 120, doi:10.1007/s00190-006-0093-1.

Coisson, P., S. M. Radicella, R. Leitinger & B. Nava (2005), Topside electron density in IRI and NeQuick: Features and limitations, *J. Adv. Space Res.*, 37 (5), 937-942.

Cornely, P.-R. J., (2003), Flexible Prior Models: Three-dimensional ionospheric tomography, *Radio Sci.*, 38, 1087, doi:10.1029/2002RS002703.

Coster, A. J., E. M. Gaposchkin, L. E. Thornton, M. Buonsanto & D. Tetenbaum (1990), Comparison of GPS and Incoherent Scatter Measurements of the Total Electron Content - The Effect of the Ionosphere on Radiowave Signals and System Performance, *MIT Lincoln Laboratory Technical Report 954.*

Daniell, R. E., (1991), Parameterized real-time ionospheric specification model: PRISM version 1.0, *Tech. Rep. PL-TR-91 – 2299*, Phillips Lab., Hanscom Air Force Base, Mass.

Daniell, R. E., L. D. Brown, D. N. Anderson, M. W. Fox, P. H. Doherty, D. T. Decker, J. J. Sojka and R. W. Schunk (1995), Parameterized Ionosphere Model: A Global Ionospheric Parameterization based on First Principles Models, *Radio Sci.*, 30, pp 1499–1510.

Davies, K., (1990), Ionospheric Radio, *IEEE Electromagn. Waves Ser.*, vol. 31, P. Peregrinus, London.

Dear, R. M., & C. N. Mitchell (2006), GPS interfrequency biases and total electron content errors in ionospheric imaging over Europe, *Radio Sci.*, 41, RS6007, doi:10.1029/2005RS003269.

Dear, R. M., & C. N. Mitchell (2007), Ionospheric imaging at mid-latitudes using both GPS and ionosondes, *J. A. S. T. P.*, 69 (7), pp 817-825, ISSN 1364-6826, 10.1016/j.jastp.2006.06.001.

Dow, J. M., R. E. Neilan & C. Rizos (2009), The International GNSS Service in a changing landscape of Global Navigation Satellite Systems, *J. Geod.*, 83(7), pp 191–198, doi:10.1007/s00190-008-0300-3.

Ely, C. V., I. S. Batista & M. A. Abdu (2012), Radio occultation electron density profiles from the FORMOSAT-3/COSMIC satellites over the Brazilian region: A comparison with Digisonde data, *Adv. Spa. Res.*, 49 (11), pp 1553-1562, ISSN 0273-1177, 10.1016/j.asr.2011.12.029.

Farley, D. T., (1969), Faraday rotation measurements using incoherent scatter, *Radio Sci.*, 4(2), pp 143-152.

Farley, D. T., J. P. Dougherty & D. W. Barron (1961), A theory of incoherent scattering of radio waves by a plasma, 2. Scattering in a magnetic field, *Proc. Royal Soc. London. Series A, Mathematical and Physical Sciences*, 263 (1313), 238–258.

Foster, J. C., M. J. Buonsanto, J. M. Holt, J. A. Klobuchar, P. Fougere, W. Pakula & B. Z. Khudukon (1994). Russian-American tomography experiment. *International Journal of Imaging Systems and Technology*, *5*(2), 148-159.

Fougere, P. F., (1995), Ionospheric radio tomography using maximum entropy 1. Theory and simulation studies, *Radio Sci.*, 30(2), 429–444, doi:10.1029/94RS03164.

Fremouw, E. J., J. A. Secan & B. M. Howe (1992), Application of stochastic inverse-theory to ionospheric tomography, *Radio Sci.*, 27, 721 – 732.

Fuller-Rowell, T. J., G. H. Millward, A. D. Richmond & M. V. Codrescu (2002), Storm-

time changes in the upper atmosphere at low latitudes, *J. Atmos. Sol-Terr. Phys.*, 64(12), 1383-1391.

Galkin, I. A., & B. W. Reinisch (2008), The New ARTIST 5 for all Digisondes, *Ionosonde Network Advisory Group bulletin.*

Galkin, I. A., B. W. Reinisch, X. Huang, and D. Bilitza (2012), Assimilation of GIRO data into a real-time IRI, *Radio Sci.*, 47, RS0L07, doi:10.1029/2011RS004952.

Gaposchkin, E., & A. Coster (1993), GPS L1-L2 bias determination, *MIT Lincoln Laboratory Technical Report 971.*

Gelb, A., (1974), Applied Optimal Estimation, *MIT Press*, Cambridge, Mass.

Gordon, R., R. Bender & G. T. Herman (1970), Algebraic Reconstruction Techniques (ART) for three dimensional electron microscopy and X-ray photography, *J. Theor. Biol.*, 29, 471 – 481.

Hajj, G. A., R. Ibanezmeier, E. R. Kursinski & L. J. Romans (1994), Imaging the ionosphere with the global positioning system, Int. J. Imaging Syst. Technol., 5, 174 – 184.

Hajj, G. A., B. D. Wilson, C. Wang, X. Pi & I. G. Rosen (2004), Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter, Radio Sci., 39, RS1S05, doi:10.1029/2002RS002859.

Hargreaves, J. K., (1979), The upper atmosphere and solar-terrestrial relations - An introduction to the aerospace environment. *New York, Van Nostrand Reinhold Co., 1979. 312 p., 273-275.*

Hargreaves, J. K., (1995), The Solar-Terrestrial Environment: An Introduction to Geospace – The Science of the Terrestrial Upper Atmosphere, Ionosphere, and Magnetosphere, *Cambridge Atmos. Space Sci. Ser.*, vol. 5, Cambridge Univ. Press., New York.

Heaton, J. A. T., S. E. Pryse & L. Kersley (1995), Improved background representation, ionosonde input and independent verification in experimental ionospheric tomography, *Ann. Geophys.*, 13, 1297 – 1302.

Heelis, R. A., J. J. Sojka, M. David & R. W. Schunk (2009), Storm-time density enhancements in the middle latitude dayside ionosphere, *J. Geophys. Res.*, 114, A03315, doi:10.1029/2008JA013690.

Hounsfield, G. N., (1972), A method of and apparatus for examination of a body by radiation such as X-ray or gamma radiation, Patent 1283915, U.S. Patent and Trademark Off., Washington, D. C.

Howe, B. M., K. Runciman & J. A. Secan (1998), Tomography of the ionosphere: Four-dimensional simulations, *Radio Sci.*, 33, 109 – 128.

Hugentobler, U., S. Schaer, T. Springer, G. Beutler, H. Bock, R. Dach, D. Ineichen, L. Mervart, M. Rothacher, U. Wild, A. Wiget, E. Brockmann, G. Weber, H. Habrich & C. Boucher (2000), *CODE IGS Analysis Center Technical Report 2000.*

Hysell, D. L., F. S. Rodrigues, J. L. Chau & J. D. Huba (2008), Full profile incoherent scatter analysis at Jicamarca, *Ann. Geophys.*, 26, pp 59–75.

Immel, T.J., G. Crowley, J.M. Forbes, R.S. Nerem & E.K. Sutton (2008), Neutral Composition and Density Effects in the October-November 2003 Magnetic Storms, Midlatitude Ionospheric Dynamics and Disturbances, *Geophysical Monograph Series 181, AGU*, pp. 259-269, doi:10.1029/181GM23

Jakowski, N., & C. Mayer (2009), Capabilities of GNSS based topside reconstructions of the ionosphere, *presented at ESA 2nd SWARM International Science Meeting 24-26 June 2009*, GFZ Potsdam, Germany.

Jee, G., R.W. Schunk & L. Scherliess (2005), Comparison of IRI-2001 with TOPEX TEC measurements, *J. Atmos. Solar-Terr. Phys.* 67, 365–380.

Jee, G., A. G. Burns, W. Wang, S. C. Solomon, R. W. Schunk, L. Scherliess, D. C.

Thompson, J. J. Sojka & L. Zhu (2007), Duration of an ionospheric data assimilation initialization of a coupled thermosphere-ionosphere model, *Space Weather*, 5, S01004, doi:10.1029/2006SW000250.

Jin, H., Y. Miyoshi, H. Fujiwara, H. Shinagawa, K. Terada, N. Terada, M. Ishii, Y. Otsuka & A. Saito (2011), Vertical connection from the tropospheric activities to the ionospheric longitudinal structure simulated by a new Earth's whole atmosphere-ionosphere coupled model, *J. Geophys. Res.*, 116, A01316, doi:10.1029/2010JA015925.

Juan, J. M., A. Rius, M. Hernandez-Pajares & J. Sanz (1997), A two-layer model of the ionosphere using global positioning system data, *Geophys. Res. Lett.*, 24, 393 – 396.

Kelley, M. C., (2009), The Earth's Ionosphere: Plasma Physics & Electrodynamics (Vol. 96). *Access Online via Elsevier.*

Kersley, L., J. A. T. Heaton, S. E. Pryse & T. D. Raymund (1993), Experimental ionospheric tomography with ionosonde input and EISCAT verification, *Ann. Geophys.*, 11, 1064 – 1074.

Kinrade, J., C. N. Mitchell, N. D. Smith, Y. Ebihara, A. T. Weatherwax & G. S. Bust (2013), GPS phase scintillation associated with optical auroral emissions: First statistical results from the geographic South Pole, *J. Geophys. Res. Space Physics*, 118, 2490–2502, doi:10.1002/jgra.50214.

Kivelson, M. G., & C. T. Russell (Eds.), (1995). Introduction to space physics. *Cambridge University Press.*

Klobuchar, J. A., (1987), Ionospheric time delay algorithm for single frequency GPS users, *IEEE Trans.*, AES-23, 325 – 331.

Kuklinski, W. S., (1997), Ionospheric tomography via iterative cross-entropy minimization, *Radio Sci.*, 32, 1037 – 1049.

Kunches, J. M., & J. A. Klobuchar (2001), Eye on the Ionosphere: GPS after SA, *GPS Solutions*, 4 (3), 52-54.

Kunitsyn, V. E., E. S. Andreeva, E. D. Tereshchenko, B. Z. Khudukon & T. Nygren (1994), Investigations of the ionosphere by satellite radiotomography, *Int. J. Imaging Syst. Technol.*, 5, 112 – 127.

Kutiev, I., P. Marinov, A. Belehaki, B. Reinisch & N. Jakowski (2009), Reconstruction of topside density profile by using the topside sounder model profiler and digisonde data, *Adv. Spa. Res.*, 43 (11), pp 1683-1687, 10.1016/j.asr.2008.08.017.

Lee, C-C., & B. W. Reinisch (2006), Quiet-condition *hmF2*, *NmF2*, and B0 variations at Jicamarca and comparison with IRI-2001 during solar maximum, *J. Atmos. Terr. Phys.*, 68, pp 2138-2146.

Lin, C. H., A. D. Richmond, G. J. Bailey, J. Y. Liu, G. Lu & R. A. Heelis (2009), Neutral wind effect in producing a storm time ionospheric additional layer in the equatorial ionization anomaly region, *J. Geophys. Res.*, 114, A09306, doi:10.1029/2009JA014050.

Ma, X. F., T. Maruyama, G. Ma & T. Takeda (2005), Three-dimensional ionospheric tomography using observation data of GPS ground receivers and ionosonde by neural network, *J. Geophys. Res.*, 110, A05308, doi:10.1029/2004JA010797.

Mandrake, L., B. Wilson, C.-M. Wang, G. Hajj, A. Mannucci & X. Pi (2005), A performance evaluation of the operational Jet Propulsion Laboratory / University of Southern California Global Assimilation Ionospheric Model (JPL/USC GAIM), *J. Geophys. Res.*, 110, A12306, doi:10.1029/2005JA011170.

Mannucci, A. J., B. D. Wilson, D.N. Yuan, C. H. Ho, U. J. Lindqwister & T. F. Runge (1998), A global mapping technique for GPS-derived ionospheric total electron content measurements, *Radio Sci.*, 33(3), 565 – 582

Markkanen, M., M. Lehtinen, T. Nygren, J. Pirttila, P. Henelius, E. Vilenius, E. D. Tereshchenko & B. Z. Khudukon (1995), Bayesian approach to satellite radiotomography with applications in the Scandinavian sector, *Ann. Geophys.*, 13, 1277 – 1287.

Materassi, M., & C. N. Mitchell (2005a), A simulation study into constructing of the sample space for ionospheric imaging, *J. Atmos. Sol. Terr. Phys.*, 67, 1085 – 1091.

Materassi, M., & C. N. Mitchell (2005b), Imaging of the equatorial ionosphere, *Ann. Geophys.*, 48(3), 477 – 482.

Matsuo, T., & E. A. Araujo-Pradere (2011), Role of thermosphere-ionosphere coupling in a global ionospheric specification, *Radio Sci.*, 46, RS0D23, doi:10.1029/2010RS004576.

McNamara, L. F., C. R. Baker & D. T. Decker (2008), Accuracy of USU-GAIM specifications of *foF2* and M(3000)F2 for a worldwide distribution of ionosonde locations, *Radio Sci.* 43, RS1011, doi:10.1029/2007RS003754.

McNamara, L. F., G. J. Bishop & J. A. Welsh (2011), Assimilation of ionosonde profiles into a global ionospheric model, *Radio Sci.* 46, RS2006, doi:10.1029/2010RS004457.

Meggs, R. W., C. N. Mitchell & V. S. C. Howells (2005), Simultaneous observations of the main trough using GPS imaging and the EISCAT radar. *Ann. Geophys.,* 23 (3), 753-757.

Mitchell, C. N., D. G. Jones, L. Kersley, S. E. Pryse & I. K. Walker (1995), Imaging of field-aligned structures in the auroral ionosphere, *Ann. Geophys.*, 13, 1311 – 1319.

Mitchell, C. N., S. E. Pryse, L. Kersley & I. K. Walker (1997a), The correction for the satellite-receiver longitude difference in ionospheric tomography, *J. Atmos. Sol. Terr.* Phys., 59, 2077 – 2087.

Mitchell, C. N., L. Kersley, J. A. T. Heaton & S. E. Pryse (1997b), Determination of the vertical electron-density profile in ionospheric tomography: Experimental results, *Ann. Geophys.*, 15, 747 – 752.

Mitchell, C. N., & P. S. J. Spencer (2003), A three-dimensional time-dependent algorithm for ionospheric imaging using GPS, *Ann. Geophys.*, 46, 687 – 696.

Mitchell, C. N., P. Yin, P. S. J. Spencer & D. Pokhotelov (2008), Ionization Dynamics During Storms of the Recent Solar Maximum, *Proceedings of the Mid-latitude Ionospheric Dynamics and Disturbances Meeting, Geophysical Monograph Series 181,* doi:10.1029/181GM09

Mlynczak, M. G., F. J. Martin-Torres, D. G. Johnson, D. P. Kratz, W. A. Traub, and K. Jucks (2004), Observations of the O(3P) fine structure line at 63 mm in the upper mesosphere and lower thermosphere, *J. Geophys. Res.*, 109, A12306, doi:10.1029/2004JA010595.

Na, H., & H. Lee (1990), Resolution analysis of tomographic reconstruction of electron density profiles in the Ionosphere, *Int. J. Imaging Syst. Technol.*, 2: 209–218. doi: 10.1002/ima.1850020307

Na, H., & E. Sutton (1994), Resolution analysis of ionospheric tomography systems, *Int. J. Imaging Syst. Technol.*, 5, 169 – 173.

Nava, B., P. Coïsson & S. M. Radicella (2009), A new version of the NeQuick ionosphere electron density model, *J. Atmos. Sol.-Terr. Phys.*, 70, pp 1856-1862, doi: 10.1016/j.jastp.2008.01.015.

Nsumei, P., B. W. Reinisch, X. Huang & D. Bilitza (2012), New Vary-Chap profile of the topside ionosphere electron density distribution for use with the IRI model and the GIRO real time data, *Radio Sci.*, 47, RS0L16, doi:10.1029/2012RS004989.

Paxton, L. J., C. I. Meng, G. H. Fountain, B. S. Ogorzalek, E. H. Darlington, J. Goldsten, S. Geary, D. Kusnierkiewicz, S. C. Lee & K. Peacock (1992), Special Sensor UV Spectrographic Imager (SSUSI): An instrument description, *Instrum. Planet. Terr. Atmos. Remote Sens.*, 1745, 2 – 16.

Pezzopane, M., & C. Scotto (2004), Software for the automatic scaling of critical frequency *foF2* and MUF(3000)F2 from ionograms applied at the Ionospheric Observatory of Gibilmanna, *Ann. Geophys. Italy*, 47(6), pp 1783 – 1790.

Pezzopane, M., & C. Scotto (2007), Automatic scaling of critical frequency *foF2* and MUF(3000)F2: A comparison between Autoscala and ARTIST 4.5 on Rome data, *Radio Sci.*, 42, RS4003, doi:10.1029/2006RS003581.

Pezzopane, M., M. Pietrella, A. Pignatelli, B. Zolesi & L. R. Cander (2011), Assimilation of autoscaled data and regional and local ionospheric models as input sources for real-time 3-D International Reference Ionosphere modeling, *Radio Sci.* 46, RS5009, doi:10.1029/2011RS004697.

Pham Thi Thu, H., C. Amory-Mazaudier & M. Le Huy (2011), Time variations of the ionosphere at the northern tropical crest of ionization at Phu Thuy, Vietnam, *Ann. Geophys.,* 29, pp 197-207.

Pi, X., C. Wang, G. A. Hajj, G. Rosen, B. D. Wilson & G. J. Bailey (2003), Estimation of E x B drift using a global assimilative ionospheric model: An observation system simulation experiment, *J. Geophys. Res.*, 108(A2), 1075, doi:10.1029/2001JA009235

Pryse, S. E., & L. Kersley (1992), A preliminary experimental test of ionospheric tomography, *J. Atmos. Terr. Phys.*, 54, 1007 – 1012.

Radicella, S. M., & R. Leitinger (2001), The evolution of the DGR approach to model electron density profiles. *Adv. Spa. Res.* 27 (1), pp 35–40.

Radon, J. (1917). On determination of functions by their integral values along certain multiplicities. *Ber. der Sachische Akademie der Wissenschaften Leipzig,(Germany)*, *69*, 262-277.

Raymund, T. D., J. R. Austen, S. J. Franke, C. H. Liu, J. A. Klobuchar & J. Stalker (1990), Application of computerized tomography to the investigation of ionospheric structures, *Radio Sci.*, 25(5), 771–789, doi:10.1029/RS025i005p00771.

Raymund, T. D., S. E. Pryse, L. Kersley & J. A. T. Heaton (1993), Tomographic reconstruction of ionospheric electron density with European incoherent scatter radar verification, *Radio Sci.*, 28(5), 811–817, doi:10.1029/93RS01102.

Raymund, T. D., Y. Bresler, D. N. Anderson & R. E. Daniell (1994), Model-assisted ionospheric tomography: A new algorithm, *Radio Sci.*, 29, 1493 – 1512.

Reinisch, B.W., & X. Huang (1983), Automatic calculation of electron density profiles from digital ionograms, 3, Processing of bottomside ionograms, *Radio Sci.*, 18, pp 477-492.

Reinisch, B.W., I. A. Galkin, G. Khmyrov, A. Kozlov & D. Kitrosser (2004), Automated collection and dissemination of ionospheric data from the digisonde network, *Adv. Radio Sci.*, 2, pp 241-247.

Reinisch, B. W., X. Huang, I. A. Galkin, V. Paznukhov & A. Kozlov (2005), Recent advances in real-time analysis of ionograms and ionospheric drift measurements with Digisondes, *J. Atmos. Sol. Terr. Phys.*, 67 (12), pp 1054-1062, doi:10.1016/j.jastp.2005.01.009.

Richards, P. G., & D. G. Torr (1986), A factor of 2 reduction in theoretical F2 peak electron density due to enhanced vibrational excitation of N_2 in summer at solar maximum, *J. Geophys. Res.*, 91(A10), 11331–11336, doi:10.1029/JA091iA10p11331.

Richmond, A. D., E. C. Ridley, & R. G. Roble (1992), A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophys. Res. Lett.*,19(6), 601–604, doi:10.1029/92GL00401.

Richmond, A. D., & G. Lu (2000), Upper-atmospheric effects of magnetic storms: a brief tutorial, *J.A.S.T.P.*, 62(12) 1115-1127, doi:10.1016/S1364-6826(00)00094-8.

Rideout, W., & A. Coster (2006). Automated GPS processing for global total electron content data, *GPS Solut.*, 10(3), 219-228.

Ridley, A. J., Y. Deng & G. Toth (2006), The global ionosphere–thermosphere model, *J. Atmos. Sol-Terr. Phys.*, 68(8), 839-864.

Rishbeth, H., (1988), Basic physics of the ionosphere: a tutorial review, *J. Inst. Elec. Rad. Eng.*, 58(6) 207-223.

Rishbeth, H., & C. S. Setty (1961), The layer at sunrise, *J. Atmos. Terr. Phys.*, 20, pp 263–276.

Rishbeth, H., & O. K. Garriott (1969). Introduction to ionospheric physics. *IEEE Transactions on Image Processing*.

Rishbeth, H, and I. C. F. Mueller-Wodarg (1999), Vertical circulation and thermospheric composition: a modelling study, *Ann. Geophys.*, 17, 794-805.

Rishbeth, H., & I.C.F. Muller-Wodarg (2006), The Annual Asymmetry in the F2-layer, *Ann. Geophys.*, 24, pp 3293-3311.

Rius, A., G. Ruffini & L. Cucurull (1997), Improving the vertical resolution of ionospheric tomography with GPS occultations, *Geophys. Res. Lett.*, 24, 2291 – 2294.

Roble, R. G., E. C. Ridley, A. D. Richmond and R. E. Dickinson (1988), A coupled thermosphere/ionosphere general circulation model, *Geophys. Res. Lett.*, **15**, 1325-1328.

Roble, R. G., & E. C. Ridley (1994), A thermosphere-ionosphere-mesosphere electrodynamics general circulation model (TIME-GCM): Equinox solar cycle minimum simulations (30 – 500 km), *Geophys. Res. Lett.*, 21, 417 – 420.

Romano, V., S. Pau, M. Pezzopane, E. Zuccheretti, B. Zolesi, G. De Franceschi & S. Locatelli (2008), The electronic Space Weather upper atmosphere (eSWua) project at INGV: advancements and state of the art, *Ann. Geophys.*, 26, pp 345–351.

Rose, J. A. R., J. R. Tong, D. J. Allain & C. N. Mitchell (2011), The use of ionospheric tomography and elevation masks to reduce the overall error in single-frequency GPS timing applications, *Adv. Spa. Res.*, 47 (2), pp. 276-288, doi:10.1016/j.asr.2010.05.030.

Sardon, E., & N. Zarraoa (1997), Estimation of Total Electron Content using GPS Data – How stable are the differential satellite and receiver instrumental biases? *Radio Sci.,* 32, 1899 – 1910

Satellite Navigation & Space Weather: Understanding the Vulnerability & Building Resilience, *A. Met. Soc. Policy Workshop Report*, March 2011

Scherliess, L., R. W. Schunk, J. J. Sojka & D. C. Thompson (2004), Development of a physics-based reduced state Kalman filter for the ionosphere, *Radio Sci.*, 39, RS1S04, doi:10.1029/2002RS002797.

Schunk, R. W., J. J. Sojka & J. V. Eccles (1997), Expanded capabilities for the ionospheric forecast model, *Rep. AFRLVS-HA-TR-98 – 0001*, 1 – 142, Space Veh. Dir., Hanscom Air Force Base, Mass.

Schunk, R. W., L. Scherliess, J. J. Sojka, and D. Thompson (2004), Global Assimilation of Ionospheric Measurements (GAIM), *Radio Sci.*, 39, RS1S02, doi:10.1029/2002RS002794.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson & L. Zhu (2005a), An operational data assimilation model of the global ionosphere, pp. 512 – 581, *Ionospheric Effects Symposium Proceedings*, Natl. Tech. Info. Serv., Springfield, Va.

Schunk, R., L. Scherliess, J. J. Sojka, D. C. Thompson & L. Zhu (2005b), Ionospheric weather forecasting on the horizon, *Space Weather*, 3, S08007, doi:10.1029/2004SW000138.

Sojka, J. J., D. C. Thompson, R. W. Schunk, V. Eccles, J. J. Makela, M. C. Kelley, S. A. Gonzoles, N. Aponte & T. W. Bullett (2003), Ionospheric data assimilation: recovery of strong mid-latitudinal density gradients, *J. A. S. T. P.*, 65, pp 2087–1097.

Sojka, J. J., M. David, R. W. Schunk, and R. A. Heelis (2012), A modeling study of the longitudinal dependence of storm time midlatitude dayside total electron content enhancements, *J. Geophys. Res.*, 117, A02315, doi:10.1029/2011JA017000.

Spencer, P. S. J., & C. N. Mitchell (2007), Imaging of fast moving electron-density structures in the polar cap, *Ann. Geophys.*, 50 (3), pp 427-434.

Stankov, S. M., K. Stegen, P. Muhtarov & R. Warnant (2011), Local ionospheric electron density profile reconstruction in real time from simultaneous ground-based GNSS and ionosonde measurements, *Adv. Space Res.*, 47, pp 1172-1180, doi: 10.1016/j.asr.2010.11.039.

Sutton, E. K., & H. Na (1994), Orthogonal decomposition framework for ionospheric tomography algorithms, *Int. J. Imaging Syst. Technol.*, 5: 106–111. doi: 10.1002/ima.1850050207

Sutton, E. K., J. M. Forbes, and R. S. Nerem (2005), Global thermospheric neutral density and wind response to the severe 2003 geomagnetic storms from CHAMP accelerometer data, *J. Geophys. Res.*, 110, A09S40, doi:10.1029/2004JA010985.

Tsai, L.-C., C. H. Liu & J. Y. Huang (2010), Three-dimensional numerical ray tracing on a phenomenological ionospheric model, *Radio Sci.*, 45, RS5017, doi:10.1029/2010RS004359.

Tsurutani, B., et al. (2004), Global dayside ionospheric uplift and enhancement associated with interplanetary electric fields, *J. Geophys. Res.*, 109, A08302, doi:10.1029/2003JA010342.

Wang, C., G. Hajj, X. Pi, I. G. Rosen & B. Wilson (2004), Development of the Global Assimilative Ionospheric Model, *Radio Sci.*, 39, RS1S06, doi:10.1029/2002RS002854.

Warrington, E. M., N. Y. Zaalov, J. S. Naylor & A. J. Stocker (2012), HF propagation modeling within the polar ionosphere, *Radio Sci.*, 47, RS0L13, doi:10.1029/2011RS004909.

Whitaker, J. S., & T. M. Hamill (2002). Ensemble data assimilation without perturbed obsevations. *Mon. Weather Rev.*, 130, 1913–24.

Wilson, B. D., & A. J. Mannucci (1993), Instrumental Biases in Ionospheric Measurements Derived from GPS Data, *Proc. ION GPS'93*, Salt Lake City.

Yeh, K. C., & T. D. Raymund (1991), Limitations of ionospheric imaging by

tomography, *Radio Sci.*, 26, 1361 – 1380.

Yin, P., & C. N. Mitchell (2005), Use of radio-occultation data for ionospheric imaging during the April 2002 disturbances, *GPS Solut.*, 9, pp 156–163.

Yin, P., C. N. Mitchell & G. S. Bust (2006), Observations of the F region height redistribution in the storm-time ionosphere over Europe and the USA using GPS imaging, *Geophys. Res. Lett.*, 33, L18803, doi:10.1029/2006GL027125.

Zapfe, B. D., M. Materassi, C. N. Mitchell, & P. Spalla (2007), Imaging of the equatorial ionospheric anomaly over South America– A simulation study of total electron content, *J.A.S.T.P.* 68 (16), 1819-1833.

Zou, L., H. Rishbeth, I. C. F. Muller-Wodarg, A. D. Aylward, G. H. Millward, T. J. Fuller-Rowell, D. W. Idenden & R. J. Moffett (2000), Annual and semiannual variations in the ionospheric F2-layer, *Ann. Geophys.*, 18, pp 927-944.

Appendix – Resolution Mapping

In order to make best use of an image, it is useful to reject the parts of the image that contain few observations. Resolution mapping is a way of rejecting poorly specified parts of a MIDAS image. MIDAS finds an optimal fit of electron densities within a grid using GPS-derived slant TEC observations and sometimes other data sources. To overcome the relative lack of vertical information contained in GPS rays, the vertical profile is constrained by basis functions. This means the MIDAS algorithm solves for basis function coefficients rather than solving directly for electron densities. The algorithm solves the following equation:

$$\boldsymbol{x}_a = \arg\min \{ (\boldsymbol{b} - A\,M\,\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{b} - A\,M\,\boldsymbol{x}) + w\,\boldsymbol{x}^{\mathrm{T}} R\,\boldsymbol{x} \} \tag{39}$$

where $\boldsymbol{x}$ is the vector of basis function coefficients, $\boldsymbol{x}_a$ is the optimal solution, $\boldsymbol{b}$ is the vector of TEC observations, A is the observation operator, M models the basis functions, R is the regularisation matrix and w is the weighting of the regularisation term. It is possible for $\boldsymbol{b}$ to contain differences between pairs of observations. This is a least-squares problem with non-regularised Hessian η :

$$\eta = M^{\mathrm{T}} A^{\mathrm{T}} A\,M \tag{40}$$

A model resolution matrix Q, as described for example by *Berryman* [2000], can be defined to capture the resolution information:

$$Q = \eta_{\mathrm{reg}}^{-1}\,\eta \tag{41}$$

where η_{reg} is the regularised Hessian:

$$\eta_{\mathrm{reg}} = M^{\mathrm{T}} A^{\mathrm{T}} A\,M + w\,R \tag{42}$$

Note that:

$$\boldsymbol{x}_{\mathrm{a}} = Q\,\boldsymbol{x}' \tag{43}$$

Hence, the diagonal elements of Q describe how accurately hypothetical 'ideal' ionosphere coefficients $\boldsymbol{x}'$ can be recovered or resolved into $\boldsymbol{x}_{\mathrm{a}}$. These are here known as resolution coefficients. The resolution coefficients describe the extent to which a cell is defined by the 'information' within it, while the off-diagonal elements

of Q show where the remainder of the information has been drawn from. Regularisation is responsible for spreading information from well-defined cells to poorly-defined cells. However, it is possible to recover resolution coefficients for each basis function in each cell. It is preferable to use the first basis function to calculate resolution coefficients, since that is the dominant and most realistic profile. A threshold between zero and one is then arbitrarily chosen and all cells with a resolution coefficient below that value are rejected. The choice of threshold depends on the specific problem being solved, since a highly regularised image will naturally have lower resolution coefficients. The resolution coefficients will vary depending on the shape of the basis functions and the ray-path geometry because it is more desirable to have ray-paths near the maximum of the function.